PATENT ABSTRACTS OF JAPAN

(11)Publication number:

2002-324000

(43) Date of publication of application: 08.11.2002

(51)Int.Cl.

G06F 12/00 G06F 3/06

(21)Application number: 2002-003193

-

(22)Date of filing:

10.01.2002

(71)Applicant : HITACHI LTD

(72)Inventor: WATANABE CHIYOKUKI

YAMAMOTO AKIRA

(30)Priority

Priority number : 2001 760344

Priority date: 12.01.2001

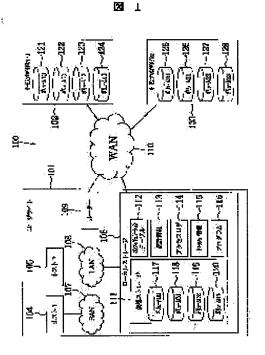
Priority country: US

(54) VIRTUAL VOLUME STORAGE

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a system for providing a data storage service.

SOLUTION: This system comprises service provider sites 102 and 103 configured to provide the data storage service, and a user site 101 coupled by a WAN(wide area network) 110 to the service provider site. The user site is provided with a local storage 106 having a virtual storage 111, the virtual storage is provided with a synchronous volume and an asynchronous volume, and the local storage is configured to immediately transmit to the data service provider data that is written in the synchronous volume, to transmit at a predetermined schedule to the service provider site data that is written in the asynchronous volume, and to read data from the service provider site if the data is not stored in the local storage.



€ 摋 4 盐 华 噩 4 (12) (19) 日本国格群庁 (JP)

特開2002-324000 (11)特許出願公開番号

(P2002-324000A)

(43)公開日 平成14年11月8日(2002.11.8)

(51) Int CL?		微別記号	F I		11.	-27-1-(物地)
	12/00	514	G06F	12/00	514E	5B065
		531			531D	5B082
		545			545A	
	3/06	304		3/06	304F	

警査請求 未請求 請求項の数17 OL 外国語出版 (全 54 頁)

(21) 田蘭番号	特膜2002—3193(P2002—3193)	(71) 出版人 000005108	80100002108
(22)出版日	平成14年1月10日(2002.1.10)		來式完在1212至作別 東京都千代田区神田駿河台四丁目6番地
(91) 阿先指小鼠 黎島	アアのリカイノロリー日本領十学に買いる	(72)発明者	滅辺 宣令 マンコナを報覧 ナコフィニアンM プロ
(32)優先日	平成13年1月12日(2001.1.12)		こくらど百米国」とうフォグーンピートラスペーン・ツイン・ジャーン・ツエッボイントズークウェイ
(33)優先權主張国	米国 (NS)		2000 日立アメリカ リサーチアンドディ
			スロップメソト部 石
		(72)発明者	山本 恭
			神奈川県川崎市麻生区王禅寺1099番地 株
			式会社日立製作所システム開発研究所内
		(74) 代理人	100091096
			弁理士 平木 祐輔
			最終買い扱く

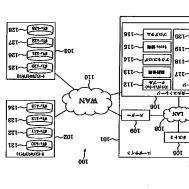
仮想ポリューム・ストレージ (54) [発明の名称]

(57) [要約]

【課題】 データ・ストレージ・サービスを提供するシ ステムを提供する。

102, 103及びサービス・プロバイダ・サイトにW AN110により結合されたユーザ・サイト101を備 ーカル・ストレージ106を備え、仮想ストレージは同 ボリュームに書き込まれたデータを所定のスケジュール データ・ストレージ・サービスを提供す るように辯成されているサービス・プロバイダ・サイト える。コーザ・サイトは仮想ストレージ111を持つロ サービス・プロバイダ・サイトに即座に送信し、非同期 でサービス・プロバイダ・サイトに送信し、データがロ ーカル・ストレージに格絶されていない場合にサービス 、プロバイダ・サイトからデータを読み込むように構成 ストレージは同期ボリューム内に書き込まれたデータを 期ボリューム及び非同期ボリュームを備え、ローカル・ **解決手段**

X



ワイド・エリア・ネットワーク (WAN) によりサービ データ・ストレージ・サービスを提供するように構成さ 「請求項1】 データ・ストレージ・サービスを提供する れているサードス・プロバイダと、 るシステムであって、

ス・プロバイダ・サイトに結合されたユーザ・サイトを カル・ストレージを備え、前記仮想ストレージは同期ボ リューム及び非同期ボリュームを備え、前記ローカル・ 備え、前記ユーザ・サイトは仮想ストレージを持つロー ストレージは前記同期ボリューム内に書き込まれたデー し、前記非同期ボリュームに書き込まれたデータを所定 のスケジュールで徴記サービス・プロベイダ・サイトに 送信し、データが前記ローカル・ストレージに格納され ていない場合に前記サーバス・プロバイダ・サイトから タを前記サービス・プロバイダ・サイトに即座に送信 データを読み込むように構成されているシステム。

【請求項2】 前記ローカル・ストレージがホストによ るように構成されていることを特徴とする請求項1記載 る当該ローカル・ストレージへのアクセス括動を記録す のシステム。

2

トレージ内のボリュームへのアクセス活動を記録するこ **前部ローカル・ストワージが前部仮想ス** とを特徴とする請求項1記載のシステム。 [請水項3]

サイトでアクセス活動を行えることを特徴とする請求項 【請求項4】 前記ローカル・ストレージが前記仮想ス トレージ内のボリュームのサブエリアでのアクセス活動 【糖水項 5】 分析のため前記サービス・プロバイダ・ を記録することを特徴とする請求項1記載のシステム。 2 記載のシステム。

データ・ストレージ・サービスを提供す 5万法であって、 [請水項6]

30

同期ボリュームと非同期ボリュームとを有する仮想スト ノージを備えるローカル・ストレージを提供するステッ

データをサービス・プロバイダ・サイトに即座に送信す 前記仮想ストレージの同期ボリューム内に書き込まれた 選択的に実行するステップであって、 るステップと、

前記仮想ストレージの非同期ボリューム内に書き込まれ

たデータを前記サービス・プロバイダ・サイトに所定の ゲータが前記ローカル・ストレージ内に格納されていな スケジュールで送信するステップと、

を読み込むステップのうち少なくとも1つを選択的に実 【請求項7】 前記ローカル・ストレージへのアクセス **こ 協合に 信記 サービス・プロベイダ・サイト かの データ 庁するステップとを含む方法。**

行するデータ・ストレージ・システム内のデータを処理 20 【請求項8】 前記サービス・プロバイダ・サイトへの

活動を記録するステップを更に含むことを特徴とする請

求項6記載の方法。

特開2002-324000

3

特幹譜状の絶囲

アクセス活動の記録を提供するステップを更に含むこと

【語水項9】 仮想ストレージの同期ボリューム内に書 き込まれたデータをサービス・プロバイダ・サイトに即 を特徴とする請求項7記載の方法。

前記仮想ストレージの非同期ボリューム内に書き込まれ たデータを前記サービス・プロバイダ・サイトに所定の スケジュールで送信するステップ、及びデータが前記仮 **都ストワージを含むローカル・ストワージ内に格徴され** ていない。場合に担記サービス・プロベイダ・サイトから データを読み込むステップのうち少なくとも1つを選択 的に実行するステップを含むデータ・ストレージ・サー ビスを提供する方法をコンピュータに実行させるための プログラムを格納する電子的に読み取り可能な媒体。 座に送信するステップ 9

書き込まれたデータをサービス・プロバイダ・サイトに 即座に送信するステップ、

スケジュールで送信するステップ、及びデータが前記仮 前記仮想ストレージの非同期ボリューム内に書き込まれ たデータを逍韶サービス・プロバイダ・サイトに所定の 想ストレージを含むローカル・ストレージ内に格納され ていない場合に前記サービス・プロバイダ・サイトから データを読み込むステップのうち少なくとも 1 つを選択 的に実行するステップを含むデータ・ストレージ・サー ビスを提供する方法をコンピュータに実行させるための 搬送液に埋め込まれたプログラム・コード。

【請求項11】 データ・ストレージ・システム内のデ ータを処理する装置であって、

ーザ・サイトであって、前記仮想ストレージは同期領域 記同期領域内に書き込まれたデータをサービス・プロバ イダ・サイトに即座に送信し、前記非同期領域に書き込 仮想ストレージを持つローカル・ストレージを備えるコ 及び非同期領域を含み、前記ローカル・ストレージは前 まれたデータを所定のスケジュールで前記サービス・ブ ロバイダ・サイトに送信し、データが前記ローカル・ス トレージに格納されていない場合に前記サービス・プロ バイダ・サイトからデータを読み込むように構成されて 【請求項12】 ローカル・ストレージへのコマンド信 いるユーザ・サイトを含むことを特徴とする装置。

仮想ストレージ内の同期領域に書き込まれたデータをサ ービス・プロバイダ・サイトに即座に送信するステッ おに応答して、 8

前記仮想ストレージ内の非同期領域に書き込まれたデー タを削配サービス・プロバイダ・サイトに所定のスケジ

データが指記ローカル・ストワージ内に格納されていな い場合に消記サービス・プロバイダ・サイトからデータ

を読み込むステップのうち少なくとも1つを選択的に実 ュールで送信するステップ

ල

データ・ストレージ・システム内のデ ータを読み込み方法であって、 [諸水項13]

ホストからローカル・ストフージへ読み込みコマンドを 受信するステップと

前記読み込みコマンドで要求されたデータのボリューム

前記ボリューム・タイプが静的である場合に、前記ロー カル・ストレージ内のローカル・ディスクからデータを ・タイプを判別するステップと、

が前記ローカル・ディスクに格納されているかどうかを **首記ボリューム・タイプがキャッシュの場合に、データ** 読み込むステップと、 調べるステップと、 前記データが前記ローカル・ディスクに格納されている 場合に、前記ローカル・ディスクからデータを読み込む

前記データが前記ローカル・ディスクに格納されていな **に格納し、前記ローカル・ディスクから前記データを読** い場合に、リモート・サービス・プロバイダ・サイトか らデータを読み込み、前記データをローカル・ディスク み込むステップと、 ステップと、

前記ローカル・ディスクからデータを読み込んだ後に、

前記読み込みコマンドに関係する統計情報を更新するス テップと、

ゲータ・ストレージ・システム内のデ 前記データを前記ホストに返すステップとを含む方法。 ータを処理する装置であって、 [請水項14]

込まれたデータを所定のスケジュールで受信し、データ がローカル・ストレージに格納されていない場合に前記 仮想ストレージ内の同期領域に書き込まれたデータを即 仮想ストレージを含む前記ローカル・ストレージにデー 座に受信し、前記仮想ストレージ内の非同期領域に書き タを送信するように構成されているサービス・プロバイ ダ・サイトを備える装置。

【請求項15】 仮想ストレージ内の同期領域に書き込 まれたデータを即座に受信するステップ、

前記仮想ストレージを含む前記ローカル・ストレージに 前記仮想ストレージ内の非同期領域に告き込まれたデー タがローカル・ストレージ内に格納されていない場合に データを送信するステップのうち 1 つを選択的に実行す るデータ・ストレージ・システム内のデータを処理する タを所定のスケジュールで受信するステップ、及びデー

データ・ストレージ・システム内にデ [請求項16]

ホストからローカル・ストレージへの割き込みコマンド **ータを書き込む方法であって、** を受信するステップと、

前記書き込みコマンドのデータのボリューム・タイプを 判別するステップと、 前記ボリューム・タイプが静的である場合に、前記ロー

カル・ストレージ内のローカル・ディスクに削記データ を書き込むステップと、 前記ボリューム・タイプがキャッシュの場合に、前記デ ータが前記ローカル・ディスクに格納されているかどう

かを調べるステップと、

前記データが前記ローカル・ディスクに格納されている **場合に、キャッシュから道部ローカル・ディスクに連記** データを書き込むステップと、

にデータ領域を割り当てて、前記データを前記キャッシ 前記データが前記ローカル・ディスクに格納されていな い場合に、前記ローカル・ディスクと前記キャッシュ内 前記データのボリューム・タイプが同期かどうかを調べ ュから前記ローカル・ディスクに書き込むステップと、 るステップと、

9

前記ボリューム・タイプが同期の場合に、前記データを リモート・サービス・プロバイダ・サイトと邙風に回想 させ、前記書き込みコマンドに関係する統計情報を更新 するステップと、

タをリモート・サービス・プロバイダ・サイトと所定の 前記ボリューム・タイプが同期でない場合に、前記デー スケジュールに基心いて同期させ、前記書き込みコマン ドに関係する統計情報を更新するステップとを含む方 23

【請求項17】 データ・ストレージ・サービスを提供 するシステムであって、

前記仮想ストレージの非同期ボリューム内に書き込まれ 仮想ストレージの同期ボリューム内に巻き込まれたデー タをサービス・プロベイダ・サイトに即座に送信、

たデータを前記サービス・プロバイダ・サイトに所定の スケジュールで送信、及びデータが前記仮想ストレージ を含むローカル・ストレージ内に格納されていない場合 ドサーブス・プロバイダ・サイト かののゲータ 調み込み のうち少なくとも 1 つを選択的に実行する手段を備える 8

[発明の詳細な説明]

発明の風する技術分野】本発明はデータ・ストレージ システムに関し、より詳細には、ユーザ・サイトに仮 **想ボリューム・データ・ストレージを備えるシステムと** 力法に関する。

[0002]

ューム、データ・バックアップなどを実施するストレー ジ関連サービスを提供する。ユーザ・サイトから、ユー 経出でこのストレージに接続する場合、ユーザは、長距 [従来の技術] 大規模なデータ・ストレージ・システム を管理することは非常に困難である。通常、データ・ス トワージ・サービス・プロバイダは、ストレージ・ボリ **デがストワージ・サービス・プロバイダのディレクトリ** る。このプロセスは、ユーザがローカル・ストレージに 難接続を利用してこのストレージに接続する必要があ 20

4

接続する場合に比べて遅延時間が長くなる。

る。プロキシ・サーバは、リモート・サーバ上で利用で ・サーバは、リモート・サーバに送られるすべての要求 を横取りし、プロキシ・サーバ側でその要求を遂行でき を提供することもできない。そのため、プロキシ・サー **べはクライアント・アプリケーション(Webブラウザ** きるアイテムをキャッシュする機能を備える。プロキシ るなどうなを聞くる。プロキシ・サーズ側たその要求を 遂行できなければ、プロキシ・サーバはその要求をリモ る。プロキシ・サーバは、リモート・サーバ又はサイト るだけであって、リモート・サーバ又はサイトへの書き べでは、資源を効果的に利用することができず、またプ ロキシ・サーバを使用した場合、ローカル・ストレージ 【0003】World Wide Web (WWW) は、広い地域で **ずは、 リーザ・サイトにプロキツ・サーバを配備するい** とができる。当業者には周知のように、プロキシ・サー など)とリモート・サーバとの間に置かれるサーバであ から送られたキャッシュ・データを読み込むことができ 込み手順をサポートしているわけではない。 さらに、ブ ロキシ・サーバでは、データの特徴に依存するサービス ストレージ・システムとして効果的な働きをする。 ユー ート・サーバに転送し、リモート・サーバに処理させ を使用した場合に比べて遅延時間が長くなる。

データ又はシステム・ファイル又はバックアップ・フ アイルといったもの)などのデータの特徴を無視してい システムを開示している。この特許では、ワイド・エリ ア・ネットワーク (WAN) 間のストレージ・システム を開示しておらず、アクセス・インタフェース・タイプ (ブロック又はファイルといったもの) 、目的 (ユーザ る。したがって、この引用で開示されているサブシステ **ヶの略陥型動的 マッピング・ゲータ・ストワージ・サン** 【0004】米国格許第5,155,835号では、 マルチレベ ムは不効率である。

ಜ

低送する。iSCSI (IP SAN)では、すべての ットワーキング・インフラにインターネット・プロトコ それにより、全体のネットワーク・コスト及び複雑 【0005】 インターネット・プロトコル小型コンピュ **一夕用周辺機器インタフェース(iSCSI)では、ネ** ル(IP)を利用し、既存のローカル・エリア・ネット 主要ネットワーキング・プロトコルをサポートできるた め、企業全体のネットワーク・アーキテクチャを統一で 度を低減できる。信頼性を保証するために、iSCSI ば、iSCSIプロトコルは、Webサイト (http://w ワーク及び/又はワイド・エリア・ネットワーク上で大 母のブロック・ストレージ (SCSI) データを来早く ではIPネットワーク用に開発された周知のネットワー ク管理ツール及びユーティリティを使用できる。たとえ セクション(http://www.ece.cmu.edu/~ips/Docs/docs. ww.ece.cmu.edu/lips/index.html) のIPストレージ

[OOO6] IETF (Internet Engineering Task Fo :ce) の作業グループはhttp://www.ietf.orgで、ネット (IETFのrfc1813) 及び共通のインターネッ 特開2002-324000 フーク・ファイル・システム (NFS) バージョン3

ト・ファイル・システム (CIFS) プロトコル (btt

p://www.cifs.org) を提供している。

ている共有ファイルにアクセスできる。NFSでは、伝 送制御プロトコル/インターネット・プロトコル(TC P/IP)の「位で実行される仮想ファイル・システム システムであり、このOSを利用すればすべてのネット ワーク・ユーザが異なる種類のコンピュータに格納され イルにアクセスできる。NFSを使用すると、ネットワ **一クに接続されているコンピュータをクライアントとし** て動作させながらリモート・ファイルにアクセスし、サ (VFS) と呼ばれるインタフェースを通じて共有ファ ーバとして動作させながらリモート・ユーザによるロー 【0007】NFSは、オープン・オペワーティング・ カル共有ファイルへのアクセスを可能にできる。

【0008】CIFSプロトコルは、数百万台のコンピ ソフトウェアをインストールすることなくファイルを共 が、ファイル及びプリンタ・アクセス用にMicrosoft Wi FSでは、すべてのアプリケーション(Webブラウザ ットフォームもコンピュータも異なるユーザでも、新規 を含む)がインターネット上のファイルを開き、共有す ュータに一度にリモート・ファイルにアクセスできるよ うにする標準を定めるものである。CIFSでは、ブラ ク(SMB)プロトコルを使用する。したがって、CI udows (発験商標) にあるサーバ・メッセージ・ブロッ 有できる。CIFSは、TCP/IP上で実行される ることができる。

トワークを接続する方法についての説明がある。WAN 接続は、非同期転送モード (ATM) 、同期光ネットワ は、ストレージ・システムとホスト・デバイスとの間の リア・ネットワーク (SAN) にワイド・エリア・ネッ [0009] Webサイトhttp://www.cnt.com及Untt b://www.san.comには、ストレージ又はストレージ・H ーク (SONET) などが考えられる。上記の引用で 略層型管理手法を実現する方法を開示していない。

[0010] 従米の方法とシステムの上述の欠陥を克服 サービス・プロバイダが信頼できるストレージ・システ ムを備えることができ、またユーザがストレージ・シス テムに素早くアクセスできるシステムと方法も必要であ **ル・ストレージ・システムをチューニングできるシステ** するシステムと力法が必要である。さらに、アクセス・ 統計データ及びユーザ・ログ・データに基へいてローカ る。さらに、アクセス・サービス・プロバイダがユーサ ムと方法も必要である。

[発明が解決しようとする課題] 本発明は、ユーザ・サ イトに信頼性の高い仮想ローカル・ストレージを使いや

20

ntm]) で説明されている。

すい形で実現することができる。本発明はさらに、コー 井向けの高速なアクセスを可能にする仮想ローカル・ス トレーンを使いやすい形で実現することができる。本統 横掛だもに、ローカル・ストレーンで追踪されるコーザ 横割デーク及びユーザ・ログ・データを使用してローカ ル・ストレージをサービス・プロバイグ側で使いやすい 形でチューンアップできるようにし、サービス・プロバ イグはこのようなチューニング・サービスについてユー ザに課金できる。 「誤題を解決するための手段】本発用の一葉施形像では、サービス・プロバイダはフィド・エリア・ネットワーグ組由でデータ・ストレージ・サービスを提供することができる。仮想ボリューム・ストレージはユーザ・サイト配幅される。ローザ・サイトのホストレージを使用することによりサービス・プロバイダ・サイトに配置されているサービス・プロバイダ・サイトに配置されているサービス・プロバイグ・プレデアクセスできる。仮越ボリューム・ストレージが利用すると、コーザはユーザ・サイトとサービス・プロバイダ・サイトの間で結合されているワイ 20ド・エリア・ネットワークに毎回でおった。ファイメ・ジャイトとサービス・プロバイダ・サイトの間で結合されているワイ 20ド・エリア・ネットワークに毎回アカスを必要がなく。他型ホネス

【0013】本発明の一実施形態では、データ・ストレ ・ジ・サービスを提供するシステムであって、データ・ ストレージ・サービスを提供するように構成されている。 により結合されたユーザ・サイトを備え、ユーザ・サイ え、仮想ストレージは同期ボリューム及び非同期ボリュ ームを備え、ローカル・ストレージは同期ボリューム内 に告き込まれたデータをサービス・プロバイダ・サイト トに送信し、データがローカル・ストレージに格納され ていない場合にサービス・プロバイダ・サイトからデー タを読み込むように構成されているシステムを広範にわ に即座に送信し、非同期ボリュームに書き込まれたデー タを呼ぶのスケジュールでキーバス・プロベイダ・サイ サービス・プロバイダ・サイト及びサービス・プロバイ ダ・サイトにワイド・エリア・ネットワーク (WAN) トは仮想ストフージを持つローカル・ストフージを備 たって提示する。

[0014] 本発明の他の実施形態では、データの特徴に基づいてローカル・ストレージ内のデータを簡単する方弦を確示する。ローカル・ストレージ内のデータを簡単する、ストレージは、データが静的データかキャッションれたデータが、同類データか非回期データかを判別する。仮想ボリューム・ストレージはおらに、ストレージ・ボリュームだけでなく、ディレクトリ、ファイル、ツリンダ、及び/又はブロック・アドレスでもデータを増

[0015]本発明の他の実権形態では、コーザ・サイト内のコーザのアクセス活動を追除する方法を提示する。コーザ・アクセス活動を記録することで、サービス 5。

・プロスイダはコーザ・アクセス活動のスターン分析やさる。この分析に基づき、サービス・プロスイダはコーザ・サイト内の仮想ボリューム・ストレージをチューニングできる。

[0016]

「発明の実施の形態」以下では、当業者が本発明を実施 し使用できるように説明し、また稀定のアプリケーショ ンとその要件の文脈において説明を行っている。実施形 他のさまざまな修正は、当業者であればたやすく難符 きるであるうし、またここで在義している一般的な原理 なびずかり一、またここで在義している一般的な原理 及びアプリケーションにも適用できる。そこで、本発明 は示されている実施形態に原たされるのではなく、ここ で開ぶされている原理、特徴および教示に一致する最も ない範囲を許容するものとする。

0012

[0017] 本発明による一実施形態では、システム100はユーザ・サイト101、さらに第1のサービス・プロバイダ・サイト102及び第2のサービス・プロバイダ・サイト103により構成される。ユーザ・サイト101の評価を以下で説明する。システム100内のサービス・プロバイダ・サイトの数は異なっていてもよ

い。第10サービス・プロバイダ・サイト102は、ストレージ・ボリューム121、122、123、及び124を備え、第2のサービス・プロバイダ・サイト103はストレージ・ボリューム125、126、127、及び128を備える。サービス・プロバイダ・サイト70のストレージ・ボリュームの数は異たっていてもよい。サービス・プロバイダ・サイト101、第10サービス・プロバイダ・サイト101、第10サービス・プロバイダ・サイト101、数5第2のサービス・プロバイダ・サイト101、数5第2のサービス・プロバイダ・サイト101、数5第2のサービス・プロバイダ・サイト101、数542のサービス・プロバイダ・サイト101の3は、WAN 110経由でデータ・ストレージ・サービスをユーザ・サイト101のユーザに辞

[0018] 当業者には周知のとおり、WANとは、通常比較的広い範囲にわたる地理的領域にまたがるコンピュータ・ネットワークのことである。WANはまた、ローカル・エリア・ネットワーク(LAN)も含む。WANに接続されているコンピュータは、多くの場合、電話システムなどの公衆網を通じて接続される。さらに、専用回線や衛星を通じて接続することもできる。現存最大のWANはインターネットである。

[0019] WAN 110では、各サイト (ユーザ・サイト101とサービス・プロバイダ・サイト102及び103) の間で通信ができる。各サイト101、102、及び103は、長い賠償で互いに隔てられていてよい。WAN 110は、通常、非同期航港モード (AT50 M)、同類光ネットワーク(SONET)、高密度数長

分割多重 (DWDM) 、又はインターネット・プロトコ れ、第2のホスト105はローカル・エリア・ネットワ サイト101を実抜することができることを当業者は 【0020】 ユーザ・サイト101では、第1のホスト 0 6 に接続されている。ホスト104及び105はロー カル・ストレージ106にアクセスする。ホスト104 及び105は、たとえば、サーバである。 ユーザ・サイ ト101内のホストの数は異なっていてもよい。 LAN 107XはLAN 108のいずれか1つ、又はSAN 107及びLAN 108の両方を含むようにユーザ 104はストレージ・エリア・ネットワーク(SAN) ーク (LAN) 108を介してローカル・ストレージ 108は、たとえば、ルータ109を介してWAN 107を介してローカル・ストワージ106に被続さ 110に接続されている。本発明の数示から、SAN ル (IP) ネットワークにより構成される。 理解するであろう。

ಜ にさらにストレージ・デバイスを追加すると、その追加 ・データ転送アーキテクチャであり、現在は、米国規格 【0021】当業者には周知のように、SANは共有ス ANによりLAN又はWAN内のすべてのサーバからす べてのストレージ・デバイスが利用可能である。SAN されたストレージ・デバイスも、さらに大きなネットワ 協会(ANSI)によって標準化されている。最も有名 -ALは、全二重データ転送を行い、転送速度は毎秒約 デバイス製造業者のコンソーシアムが開発したシリアル なファイバ・チャネル標準は、Fibre Channel Arbitrat や非常に高い帯域幅を必要とするその他の周辺機器用に 設計されている。光ファイバでデバイスを接続するFC [0022] SAN 107は、たとえば、ファイバ・ チャネル又は小型コンピュータ用周辺機器インタフェー に、ファイバ・チャネルはコンピュータ及び大容量記憶 ed Loop (FC-AL)で、新しい大容量記憶デバイス ーク内の任意のサーバからアクセスできるようになる。 ス(SCSI)で構成できる。当業者には周知のよう トレージ・デバイスの高速サブネットワークであり、 100メガバイト (MBps) である。

ともできる。

【の023】また当業者には周始のように、SCSIは 種々のコンピュータシステムで周辺装置をコンピュータ に接続するために使用されているパラレル・インタフェ 一ス標準である。SCSIインタフェースは、標準装備 のシリアル・ポート及びパラレル・ボートに比べて高速 なデータ転送速度(最大約80MBps)を利用でき

[0024]また当業者には同的のように、LANとは、通常比較的狭い地域に置かれるコンピュータ・ネットワークのことである。ほとんどのLANは、単一の離物又は建物群に制限されている。ほとんどのLANは、ワークステーション及びパーンナル・コンピュータを検続する。LAN内の各ノード(個々のコンピュータを検続する。LAN内の各ノード(個々のコンピュータを検

特開2002-324000

9

は、個々に中央処理装置(CPU)を備え、プログラムを実行するが、LAN上の任意の場所にあるデータ及びデバイスにアクセスすることも可能である。したがって、多くのユーザが、レーザ・ブリンタなどの高価なデバイスをデータとともに共有できる。ユーザはちに、セムできる。LANにはいるいろな種類のものがある。よっとにはイーサネット、登録商標)が最も一般ある。LANは非常に適進にデータを送信することができ、電話回線でデータを送信するよりもかなり高速である。しかし、LANの配袋距離は開展されており、また単一のLANに複銭できるコンピュータの台数にも問題

10025] 図1では、LAN 108をたとえばイーサネット(登録商標)として構成している。第1のホスト104は、たとえば、ブロック(SCS1)インタフェース(1/F)を使用してローカル・ストレージ106にアクセスする。第2のホスト105は、たとえば、20フィル(NFS又はCIFS)1/Fを使用してローカル・ストレージ106にアクセスする。第2のホスト105は、たとえば、プロック(iSCS1)1/Fを使用してローカル・ストレージ106にアクセスする。

ファイバ・チャネル、又は独自ベンダ・プロトコルを使 でiSCSIXはファイバ・チャネル、SONET上で 用できる。ローカル・ストレージ106は、仮想ストレ (図4も参照)、統計情報113 (図5も参照)、アク を含む。仮想ストレージ111内のボリュームの数は異 【0026】ローカル・ストレージ106及びサービス ・プロバイダ・サイト102及び103では、A TM上 ージ111、コンフィギュレーション・テーブル112 セス・ログ114 (図6も参照) 、キャッシュ情報11 5、及びプログラム116を含む。コンポーネント11 2、113、114、115、及び116ではローカル ・ストレージ106をキャッシュとして動作させること 11にアクセスできる。仮想ストレージ111は、いく つかのボリューム117、118、119、及び120 なっていてもよい。これらのボリューム117~120 の管理は、ローカル・ストレージ106が行う。ローカ ル・ストレージ106及びサービス・プロバイダ・サイ ト102及び103は、仮想ボリューム・ストレージ1 ができる。ホスト104及び105は仮想ストレージ1 11の作成のため共同作業を行う。 49

[0027] コーザは、コーザ・サイト101に仮想ボリューム・ストレージ111 (ローカル・ストレージ106内の) を配循することができる。仮想ボリューム・ストレージ111が4ーザ・サイト101に配備されると、ホスト104及び/又はホスト105のコーザは仮8がメコーム・ストレージ111を使用することによ

ューム・ストレージ111は、たとえば以下で説明する を備える。仮想ボリューム・ストレージ111は、ユー ではさらに、サービス・プロバイダがユーザ向けに高速 -ム121~124及び/火はボリューム125~12 1 1 は通常、サービス・プロバイダ・ストレージ・ボリ (NFS、CIFS) などのいくつかのインタフェース ザがWAN 110に毎回アクセスしなくてよいように すると都合がよい。仮想ボリューム・ストレージ111 で信頼性の高いストレージ・システムを提供することも り、巨大なサービス・プロバイダ・ストレージ(ボリュ ム125~128) に比べてサイズが小さい。仮想ボリ ューム (ボリューム121~124及び/又はボリュー 8)にアクセスできる。仮想ボリューム・ストレージ 1 ようなブロック1/F (SCS1)、ファイル1/F

トレージ106は、マイクロプロセッサ201 (たとえ ローカル (内部) バス202、SAN 107に結合さ 実施形態の詳細を示すブロック図である。ローカル・ス 1)、LAN 108に結合されているLAN 1/F 【0028】図2は、ローカル・ストレージ106の一 れているSANインタフェース (1/F) 203 (図 ば、Intel CorporationやMotorola Corporation製)

204 (図1)、WAN 110に結合されているW メモリ207、ディスク・コントローラ208、及びロ **一カル・ディスク209を備える。メモリ207は、コ** 3、アクセス・ログ114、キャッシュ情報115、及 0.1は、ローカル・ストレージ106内のすべてのリン ぴプログラム116を格納する。マイクロプロセッサ2 ソフィギュレーション・テーブル112、統計信報11 AN 1/F 205 (図1)、メモリ1/F206、 一スを制御し、プログラム116を使用してローカル・ ストレージ106内のすべてのプロシージャを実行す

イスク209に格塗されている。ローカル・ストレージ [0029] 図2は、ローカル・ストレージ106内の 図2に示されている情報及びプログラムはローカル・デ 106のブート手順で、これらのデータ及びプログラム はローカル・ディスク209からメモリ207に移動さ 209のスナップショットを示している。 当業者には固 知のとおり、スナップショットとは、実際のストレージ プロシージャを実行しているときのローカル・ディスク の全ボリューム又はボリュームの一部のコピーである。

ナッシュ・データ212は、LRU (Least Recently U 50 して動作する。SCSIサーバ216は、SCSIター [0030] キャッシュ情報115は、キャッシュ・デ スク209内のキャッシュされたデータ212のディレ クトリ(コンフィギュレーション情報)である。このキ イレクトリ210、保留データ・リスト211、及びキ ヤッシュ・データ212で構成される。 キャッシュ・デ イレクトリ210は、メモリ207及びローカル・ディ

等限2002-324000

來再びアクセスする可能性がないことで、LRU管理の 列外となるケースがある。このような例外の1つに順次 アクセスがあり、ストレージ・アドレスが逐次的にアク sed) 管理法で管理することができる。しかし、近い将

ストレージ106に保存されている保留データのリスト である。保留データ・リスト211は、保留データへの 【0031】保留データ・リスト211は、ローカル・ ポインタ、サービス・プロバイダ・サイトへのポイン

タ、及び同期期間などの各保留データの情報を含む。同 期期間は、データの特徴により定義される。データが重 要でない場合、このデータは所定の期間、ローカル・ス は、たとえば、1分、1時間、1日、1週間、又は1カ トレージ106内に格納することができる。 この期間

ーション・テーブル112 (図4)を使用してデータを 月とすることができる。データが重要な場合、データが 仮想ストレージ1116格納された後、コンフィギュレ **即座に (回期して) サービス・プロバイダ・サイト (た** とえば、サイト102)に格納すべきである。たとえ ば、同期データはボリューム00 117に格納され

00 で観別される。ボリューム00 117 に格納され ているデータは同期データなので、データはサービス・ プロベイダ・サイト102 (リモート・サイト1)のボ リューム10 121 (リモート1D 10) に即座に る。図4で、ボリューム00 117は識別番号ID 格納される。

【0032】この同期アクセス機能を使用すると、信頼 生の高いデータ・ストレージを構築できる。ただし、ユ ーザ・サイト101からサービス・プロバイダ・サイト 102及び/又は103にアクセスするのに時間がかか 必要なアクセス時間は、システム100で実行される特 るため、同期アクセスにはかなりの時間が必要である。 帝の才ペアーションによって異なる。

タは、キャッシュ・データ212と呼ばれる。キャッシ [0033] ホスト104及び/又は105からアクセ スされたデータはメモリ207に格納される。このデー ュ・データ212は、ホスト104及び/又はホスト1 05のユーザによって頻繁にアクセスされる一時的デー タである。

【0034】一実施形態では、プログラム116は、サ プログラム213は、NFSサーバ214、CIFSサ ーバ・プログラム213、シンクロナイザ217、キャ ッシュ制御218、スケジューラ220、及びデバイス ・ドライバ219を含む。サーバ・プログラム213で NFSサーバ214は、従来のNFSサーバとして動作 する。CIFSサーバ215、従来のCIFSサーバと ま、ローカル・ストレージ106及びホスト104 (及 び/又はホスト105)が互いに通信できる。サーバ・ -バ215、及びSCSIサーバ216で構成される。

<u>@</u>

7は、保留データ・リスト211を使用して、ローカル どのすべてのプロシージャをスケジュールする。デバイ ス・ドライパ219は、SAN I/F 203、LA 1/F 206、及びディスク・コントローラ208な **ヤッシュ制御218、及びデバイス・ドライバ219**な N I/F 204, WAN I/F 205, メモリ どのローカル・ストレージ106内のすべてのデバイス 02及び/又は103の同期をとる。キャッシュ制御2 メモリ207とローカル・ディスク209内のキャッシ ゲット・デバイスとして動作する。シンクロナイザ21 ・ストレージ106とサービス・プロバイダ・サイト1 18は、キャッシュ・ディレクトリ210を使用して、 ユ・データ212を制御する。スケジューラ220は、 サーバ・プログラム213、シンクロナイザ217、

じてデータの特徴を取り扱う。以下でさらに詳しく説明 ストレージ106の管理は、以下でさらに詳しく説明し ているように、データの特徴を使用することに基づいて いる。仮想ボリューム・ストレージ111では、データ が静的かキャッシュされているか、同期か非同期かに応 り、ファイル、シリンダ、及び/又はブロック・アドレ スでもデータの特徴を取り扱える。ユーザ・サイト10 1では、仮想ストレージ111に4種類の仮想ボリュー ム (ボリューム00 117、ボリューム01 11 8、ボリューム02 119、及びボリューム03 1 (1) 静的同期ボリューム、(2)静的非同期ボリコー (4) キャッシュされた非同期ボリュームについて次に 20)がある。これらの仮想ボリュームの特徴である、 レイアウトの一倒を示すプロック図である。ローカル 【0035】図3は、本発明の実施形態によるデータ しているように、仮想ボリューム・ストレージ111 は、ストワージ・ボリュームだけでなく、ディレクト ム、(3) キャッシュされた同期ボリューム、及び

て実際のストレージ・ボリュームであるボリューム10 ム、ボリューム00 117は、高速な読み込みアクセ 121と同じサイズを占有する。ホスト104及び/ 又は105 (ユーザ・サイト101の) は、ボリューム 121からは読み込まない。アクセスされたデータがキ **場合、ローカル・ストレージ106(図2)はローカル** ・ディスク209 (図2) にアクセスせず、メモリ20 1 にアクセスするだけである。ホスト1 0 4及び/又は [0036] (1) 静的同期ボリューム (ボリューム0 0 117):ボリューム00 117は静的同期ボリ は、第1のサービス・プロバイダ・サイト102におい ヤッシュ・データ・キャッシュ212(図2)内にある ュームである。したがって、ポリューム00 117 00 117からデータを読み込み、ボリューム10 105は、ボリューム00 117とボリューム10 121にデータを書き込む。この静的非同期ボリュー

高いストレージ・システムを構築できる。

20

等開2002-324000 【0037】 (2) 静的非同期ボリューム(ボリューム 118) : ボリューム01 118は静的非同期 18は、第1のサービス・プロバイダ・サイト102に おいて実際のストレージ・ボリュームであるボリューム 11 122と同じサイズを占有する。ホスト104及 び/又は105は、ボリューム01 118からだけデ ータを読み込み、ボリューム11 122からはデータ データ212内にある場合、ローカル・ストレージ10 6はローカル・ディスク209にアクセスせず、メモリ 又は105は、ボリューム01 118にデータを書き この保留データは、以下で説明するように、所定のスケ 207にアクセスするだけである。ホスト104及び/ 込み、保留データ・リスト211 (図2) に登録する。 を読み込まない。 アクセスされたデータがキャッシュ ポリュームである。したがって、ポリューム001 スと信頼性の高いストレージ・システムを実現する。 ジュールで同期処理される。

【0038】この保留データは、バックグラウンド・ジ し、WAN 110のトラフィックを低減することがで 18は、ボリュームの高速な読み書きを行うが、信頼度 ョブでローカル・ストレージ106(図2)内に圧縮 [0039] (3) キャッシュされた同期ボリューム きる。この静的非同期ボリューム、ボリューム01 は同期ボリュームの場合と同じではないことがある。

作する。したがって、ホスト104及び/又は105に イスク209にアクセスせず、メモリ207にアクセス (ボリューム02 119):ボリューム02 119 て、ボリューム02 119は、第1のサービス・プロ パイダ・サイト102において実際のストレージ・ボリ ュームであるボリューム12 123と同じサイズを占 有する。ボリューム02 119はキャッシュとして動 よってアクセスされるデータがボリューム02 119 内にない。アクセスされたデータがローカル・ディスク 209 (図2) 内にない場合、ローカル・ストレージ1 06 (図2) は郑1のサービス・プロバイダ・サイト1 02からデータを読み込んで、読み込んだデータをロー 106はメモリ207を介してローカル・ディスク20 アクセスされたデータがキャッシュ・データ212内に ある場合、ローカル・ストレージ106はローカル・デ するだけである。ホスト104及び/又は105は、ボ を書き込む。このキャッシュされた同期ボリューム、ボ リューム02 119は都合よく、ローカル・ストレー ジ106のサイズを減らすことができる。この仮想ボリ ューム、ボリューム02 119ではさらに、信頼性の **メトワー沙106内に入りた後、ローカガ・メトワーツ** リューム02 119とボリューム12123にデータ 9からデータをホスト104及び/又は105に送る。 カル・ディスク209に書き込む。データがローカル はキャッシュされた同期ボリュームである。 したがっ

イトで提供しているサービスも示している。第1のサー 30 [0041] 図3はさらに、サービス・プロバイダ・サ ビス・プロバイダ・サイト102のボリューム11 1 22は、第2のサービス・プロバイダ・サイト103内 のボリューム21 126に複製される。ボリューム1 2 123は、第2のサービス・プロバイダ・サイト1 03内のボリューム22 127に複製される。ボリュ ーム13 124は、第2のサービス・プロバイダ・サ ボリューム・データのこの複製 (リモート・コピー13 たとえば、従来のリモート・ミラーリング技術を使用し てもボリューム12 123のデータのバックアップを サービスにより、ユーザはデータ・バックアップの作業 サイト102では、ユーザ・サイト101の機能がなく て複製できる。さらに、第1のサービス・プロバイダ・ イト103内のボリューム23 128に複製される。 実行できる(矢印132を参照)。 このバックアップ・ 0) は、障害回復方法となる。ボリューム・データは、 負荷を軽減できる。

【0042】以下で説明しているように、仮想ポリューム・ストレージ111では、ポスト104及び/又は105からのアクセス活動に基づいてトレース・データを作成できる。仮想ポリューム・ストレージ111は、サ 50

0を参照)を持ち、ポインタ (「ptr」) はログP t r

[0043] 図4は、本窓町の実施形態によるコンフィギュレーション・テーブル1120一倒を示すプロックのである。コンフィギュレーション・テーブル112は、パラメータ「ID」、「インタフェース」、「サブエリア」、「リモート・サイト」、「リモートID」、「ボリューム・タイプ」、「サービス」(「ペックアップ」、「分析」)、「サイズ」(「合計」、「使用済み」、「空き」)、及び「ボリシー」で構成される。

ェース」は、SCSI、NFS、及びCIFSなどの仮 エリア」は、仮想ボリューム内のサブエリアの個数を示 す。サブエリアは、ブロック・アドレス、シリンダ、フ プロバイダ・サイト102のボリューム10 121は 「ID」は、ローカル・ストレージ106内の仮想ボリ ュームのボリューム識別を示す。たとえば、ボリューム 00 117 (図3)のIDは00である。「インタフ 想ボリューム・インタフェース・タイプを示す。「サブ とができる。「ボリューム・タイプ」は、サブエリア間 で異なることがある。この場合、図4の実例で、各サブ ボリューム00 117 (図3) (IDは00) は上述 のように同じ静的同期ボリューム・タイプである。「リ モート・サイト」は、リモート・サイトの簔別子を示す (たとえば、サービス・プロバイダ・サイト102の織 別子)。これは、httpアドレスのIPアドレスでも よい。「リモートID」は、サービス・プロバイダ・サ イト内のボリュームIDを示す。たとえば、サービス・ ァイル、及び/又はディレクトリのあるエリアとするこ エリアはボリューム・タイプが同じである。たとえば、 リモートID 10を持つ。「ボリューム・タイプ」

ッシュされた非同級ボリューム、ボリューム03 120は都合よく、ローカル・ストレージ106のサイズを減らすことができるが、信機性は同級ボリュームと同じ

にならない場合がある。

は、静的な又はキャッシュされた、同期又は非同期など のボリュームのタイプを示す。「サービス」は、バック アップ・サービスや分析サービスなどのサービス・プロ バイダが提供するサービスを示す。サービス・プロバイ ダは、分析サービスを設定し、統計情報(1つをス・ (国2) へのポインをを設定し、統計情報(フクセス・ (国2) へのポインをを設定し、統計情報(フクセス・ は、分析サービスを設定し、統計情報(フクセス・ プレビスが提供される場合、図4の側に示されているよ カービスが提供される場合、図4の側に示されているよ プレビスが提供される場合、図4の側に示されているよ プレビスが提供される場合、図4の側に示されているよ プレビスが指でクション・デーブル112のボリューム02 119(1D02)は、分析セクションにボインタ1 で着つ(行400を参照)。(ボリューム02 119(1D02)は、分析セクションにボインタ1 の)1D 02に対するこのボインタは、図5内の統 評価機113の中の1D 02を指している。1D 0 2は、サブエリア(ディレクトリ)「fusral(行50

17

9

セクション内にある。このポインタ (ptr) は、図6の例のボリューム02 119に対し存成されたアクセス・ログであるアクセス・ログ114を指している。

[0044] さらに、図5の例では、(ボリューム02119の) ID 02はディレクトリ「varb」(行505参照) を含み、「Null」値 (ポインタなし) がログP trセクション内にある。さらに、ID 02は、ディレクトリ「varc」(行510を参照) を含み、Null値 (ポインタなし) がログP trセクション内にある。 たがって、ディレクトリ/varb及び/varc にフクトコーグを指すポインタがないため、これらのディレクトリに対し認連するアクセス・ログ144が作成されたい。

20 ある。パラメータ「ポリシー」は、同期のスケジュール 06内にない場合、合計サイズのみ判明している必要が (IDは02) は、「ポリシー」パラメータで「週」値 ズ(「合計」)を示し、また使用済みサイズ(「使用済 17 (10は00)は、図4で「ポリシー」パラメータ に「Null]値が設定されているが、このNull値 は同期スケジュールがボリューム00 117に対し設 を設定しており、この値は、ボリューム02 119内 とえば、サイト102)のデータと同期処理されている [0046] コンフィギュレーション・テーブル112 【0045】 コンフィギュレーション・デーブル112 で、「サイズ」は、ギガバイト単位で合計メモリ・サイ み」)及び空きサイズ(「空き」)を示している。ボリ ュームがファイク・システムのローカル・ストワージ」 を示す。たとえば、行410では、ボリューム00 1 のデータが週に1度サービス・プロバイダ・サイト (た り、この値は、ボリューム03 120内のデータが毎 日サービス・プロバイダ・サイト (たとえば、サイト1 定されていないことを示す。ボリューム02 119 は、「ポリシー」パラメータで「日」値を設定してお ことを示す。ボリューム03 120 (IDは03) 02)のデータと同期処理されていることを示す。

[0047] 図5は、本発明の実施形態による統計情報 113の一例を示すプロック図である。ユーボスはサービス・プロバイダがコンフィギュレーション・テーブル 112 (図4) で「サービス分析」パラメータを「P・「」に設定した場合、次のことが可能になる。ローカル・ストレージ106のユーザ統計アクセス情報を収集する。「統計情報」は、「ID」、はアエリガ、「IO/s」及び「MB/s」(「平均、域大部を込み」、「下均、域大部を込み」、「下均、域大部を込み」、「下均」、域大部を込み」、「下均」、域大部を込み」、「下均」、域大部を込み」、「下均」、域大部を込み」、「下均」、域大部のプログP+r」で構成される。パラメータ「ID」

特開2002-324000

18 及び「サブエリア」(図5) は、コンフィギュレーション・テーブル112 (図4) の「1D」及び「サブエリア」と同じ意味を持つ。「1O~5」は、1秒をたりの読み込み及び母き込みコマンドなど、1秒あたりのホスト/ユーザ入力又は出力活動の個数を示す。ポリューム又はポリューム・ディレクトリ内の読み込み又は書き込みアクセスは、知られている適当な方法で記録できる。

100431 15ツ7年11, ローカル・イトレーンの ヒット華を示す。このヒット華は、たとえば、ローカル ・ディスク209 (図2) だけのものである。ヒット華 は、次の式(1)で定義される。 (1) ヒットキ=100% (#ローカル・ディスク糖み

込み/4全部か込み) ただし、「甘ローカル・ディスク語み込み」は、ホスト 別のローカル・ディスク209内の親み込み回敷であり、「年全部み込み」は、ホスト別の合計部み込み回敷でり、「「年金部み込み」は、ホスト別の合計部み込み回敷 [0050]統計情報113は、サービス・プロバイダ・サイト102及び/又は103に定期的に送られる。サービス・プロバイダでは、データの特徴を分析し、この分析結果からサービス・プロバイダはユーザ・サイト101のホストのユーザに適切な解決策を整示すること

ができる。 【0051】上述のように、ログPtrは、アクセス・ログ114~のがインタを示している。サービス・プロバイダは、ユーザ・アクセス・パターンの評蓄を知りたい場合、このポインタを作成されたアクセス・ログ11 4 に設定する。
10052] 図6は、本発明の実施形態によるアクセス・ログ114の一例を示すプロック図である。サービス・プロバイダは、コーザのアクセス・バターンの詳細を分析するときに、アクセス・ログ114を作成し、ログリア (ログクセス・ログ114を作成し、ログレージ106は、ユーザ別にある)のボインをとしてファス・ログ114に設定する。ローカル・ストレージ106は、ユーザ別になってファイ)、「ファイル1D」、「アドレス」、及び「サイズ」で構成される。 目付」は、ユーザがアクセスした日付を示す。「時間」は、ユーザがアクセスした日付を示す。「時間」は、コーザがアクセスした日付を示す。「時間」は、コーザがアクセスした日付を示す。「時間」は、コーザがアクセスした日付を示す。「時間」は、コーザがアクセスした日付を示す。「時間」は、コーナがアルを

とにより設定される。各ストレージ・サブシステムは閲

連するコンフィギュレーション・アーブルを持つ。

1 2内のコンレイギュレーション・ゲータを設定するい

のパラメータは、コンフィギュレーション・テーブル」

「日付」は、ユーナがアクセスした日付を示す。「時 刻」は、アクセスの時刻を示す。「コマンド」は、アク セスのコマンド・タイプ(たとえば、読み込み又は書き 込み)を示す。「ファイルID」は、このコマンドでア 50 クセスしたファイルの鍵別を示す。アクセス・ログおS

こSI用であれば、「ファイルID」はNullとする こともできる。「アドレス」及び「サイズ」は、アクセ ス・アドレス及びサイズを示す。

[0053] 読み込みプロセス

みコマンドを分析して、読み込みコマンドをサーバ・プ I/F 203XはLAN I/F 204)を介して (ローカル・ストレージ106) はキャッシュとして動 ストレージ106にアクセスするが、これは、ローカル ・ストレージ106の空き容量が大きいからである。ま トコルの読み込みコマンドを発行する。ローカル・スト I/F 203 (図2)を介して読み込みコマンドを受 信する。デバイス・ドライバ219 (図2) は、ホスト 一ラ220に入れる。次にスケジューラ220は読み込 ログラム213内において適切なサーバ (NFSサーバ 214, CIFSサーバ215, XはSCSIサーバ2 る。各サーバ・プログラム213は、鷲み込みコマンド 2) 内にあるかどうかを調べる。 すべてのデータ (読み 2内にある場合は、ローカル・ストレージ106は内部 競み込みプロシージャでは、キャッシュされたデバイス 作する。ホスト104及び/又は105は、ローカル・ FS、又はiSCSIプロトコルの髋み込みコマンドを 発行するか、又はSAN 107を介してSCSIプロ からのこのコマンドを処理し、このコマンドをスケジュ 16)に入れて読み込みコマンドを処理することができ 込みコマンドで要求された)がキャッシュ・データ21 は、LAN 108 (図1)を介してNFS、CI ず、ホスト (たとえば、ホスト104又はホスト10 で要求されたデータがキャッシュ・データ212 (図 レージ106は、LAN I/F 204XはSAN バス202とネットワーク・インタフェース (SAN

ら読み込みコマンドを受け取った後、ローカル・ストレ データを要求側ホスト (ホスト104又は105) に返 す。他方、データ(読み込みコマンドで要求された)の ら、又はサービス・プロバイダ・サイト102又は10 06がホスト (たとえば、ホスト104又は105) か ージ106はデータのボリューム・タイプを調べる(ス テップ701)。スケジューラ220は、ボリューム1 Dをチェックすることによりコンフィギュレーション・ テーブル112内のボリューム・タイプ (たとえば、静 る (取り扱う) サーバ・プログラム (サーバ214、2 【0054】図1は、本発明の実施形態によるこの読み 2) は、読み込みコマンドを分析して、データのボリュ **ーム・タイプを判別し、また読み込みコマンドを処理す** 3からキャッシュ・データ212に移動すべきである。 込みプロセスの流れ図である。ローカル・ストレージ 1 全部又は一部がキャッシュ・データ212内にない場 合、要求されたデータはローカル・ディスク209か 的タイプ)を調べる。特に、スケジューラ220(図 15、又は216)を決定する。スケジューラ220

タイプと適切なサーバ・プログラムを調べる。ボリュー ム・タイプが静的ボリュームの場合、ローカル・ストレ **一ジ106はステップ702、703、及び704を飛** ばし、以下で説明するステップ705に進む。 ボリュー カル・ストレージ106はデータ (読み込みコマンドで ム・タイプがキャッシュされたボリュームの場合、ロー 要求した) がキャッシュ・データ212内に格納されて を見て歌み込みコマンドを処理するためのボリューム・ いるかいないかを調べる (ステップ702)。

【0055】ローカル・ストレージ106は、キャッシ 4、215、又は216)は、読み込みコマンドで要求 したデータがローカル・ディスク209内にあるかどう した)がローカル・ディスク209に格納されている場 合、ローカル・ストレージ106はステップ704を飛 ばして、以下で説明するステップ705に進む。データ 特に、サーバ・プログラム213の適切なサーバ (21 かを調べる。すべてのデータ(読み込みコャンドで要求 (読み込みコマンドで要求した) がローカル・ディスク 209内にない場合、ローカル・ストレージ106はス コ・ディレクトリ210を띏べる(ステップ703)。 テップ704を実行する。

込みコマンドで要求した)を読み込む。特に、キャッシ 【0056】ステップ104で、ローカル・ストレージ 106はローカル・ディスク209内でデータ領域を割 り当て、サービス・プロバイダ・サイト(たとえば、サ ービス・プロバイダ・サイト102) からデータ (幌み **4 制御218では、コンフィギュワーション・アーブル** 112 (図4) を使用してデータが取り出されるサービ ス・プロバイダ・サイトを決定する。キャッシュ制御2 18では、サービス・プロバイダ・サイトからデータを 説み込み、データをローカル・ディスク209に格納す

更新する (ステップ106)。 特に、ローカル・ストレ 【0057】データをローカル・ディスク209に格納 スク209からキャッシュ・データ212ヘデータを移 助する(書き込む) (ステップ705)。特に、キャッ シュ制御218は、ローカル・ディスク209からキャ カル・ストレージ106は、統計情報113 (図5)を 一ジ106のスケジューラ220は、統計情報113を 更新する。統計情報 1 1 3 内のログ・ポインタ (ログP 1 г) (図5) が設定されている場合、ローカル・スト レージ106はこの読み込みコマンドのログ・データを 5220又はデバイス・ドライバ219は、統計情報1 | 3とアクセス・ログ114を作成する。通常、統計情 した後、ローカル・ストレージ106はローカル・ディ アクセス・ログ114 (図6) に追加する。スケジュー 服113とアクセス・ログ114の作成においてはスケ **ッシュ・データ212にデータを移動する。続いてロー** ジューラ220が用いられる。

【0058】その後、データ(読み込みコマンドで要求

20

は、コンフィギュレーション・ゲーブル112(図4)

(12)

存開2002-324000

ゲラム (214、215、Xは216) に移動し、適切 した)はキャッシュ・データ212から、鷲み込みコマ **ッシュ・データ212からデータを適切なサーバ・プロ** 5)に移動される。特に、キャッシュ制御218はキャ なサーバ・プログラム (214、214、又は216) がデータを、読み込みコマンドを送ったホストに返す。 ンドを送ったホスト (たとえばホスト104义は10 [0059] <u>書き込みプロセス</u>

(ローカル・ストレージ106) はキャッシュとして慰 203を介して書き込みコマンドを受信する。デバイス ・データ212内の)を割り当てる。 すべてのデータ領 はiSCSIプロトコルの書き込みコマンドを発行する か、又はSAN 107を介してSCSIプロトコルの あるかどうかを調べる。すべてのデータ(者き込みコマ ローカル・ストレージ106は内部バス202とネット ワーク・インタフェース (SAN I/F 203又は た空き領域にデータを上書きする。他方、(書き込みコ てある領域(ローカル・ディスク209及びキャッシュ **割き込みプロシージャでは、キャッシュされたデバイス** 作する。ホスト104及び/又は105は、ローカル・ ストレージ106にアクセスするが、これは、ローカル 5) は、LAN 108を介してNFS、CIFS、又 ドライバ219 (図2)は、ホストからのこの読み込 みコマンドを処理し、このコマンドをスケジューラ22 0に人れる。次にスケジューラ220は書き込みコマン ドを分析して、書き込みコマンドをサーバ・プログラム 入れる。各サーバ・プログラム213は、書き込みコマ ンドのデータがキャッシュ・データ212 (図2) 内に LAN I/F 204)を介してホストからデータを ディスク209及びキャッシュ212内の割り当てられ キャッシュ・データ212内にないデータの残りについ 域の割り当てが済んだら、ローカル・ストレージ106 はその割り当てられた領域内にデータを格納する。すべ てのデータをキャッシュ・データ212内に格納する場 合、ローカル・ストレージ106はローカル・ディスク 書き込みコマンドを発行する。ローカル・ストレージ 1 ンドの) がキャッシュ・データ212内にある場合は、 受信する。ローカル・ストレージ106は、ローカル・ マンドの)データの全部又は一部がキャッシュ・データ 2 1 3 内において適切なサーバ (NFSサーバ2 1 4、 06t, LAN I/F 204XtSAN I/F 212内にない場合、ローカル・ストレージ106は、 ず、ホスト (たとえば、ホスト104又はホスト10 ・ストレージ106の空き容量が大きいからである。 CIFS#-1215, XISCSI#-1216)

20 プロセスの流れ図である。ローカル・ストレージ106 【0060】図8は、本発明の実施形態による書き込み

ラムを調べる。ボリューム・タイプが静的ボリュームの がホストから書き込みコマンドを受け取った後、ローカ **バ・ストワージ 106 は勘き込みコセンドのデータのボ** スケジューラ220(図2)は、書き込みコマンドを分 折して、データのボリューム・タイプを判別し、また懿 み込みコマンドを処理する(取り扱う)サーバ・プログ テーブル112(図4)を見て書き込みコマンドを処理 するためのボリューム・タイプと適切なサーバ・プログ 803、及び804を飛ばし、以下で説明するステップ 805に進む。ボリューム・タイプがキャッシュされた **ポリュームの基合、ローカル・ストレージ106はデー** タ (書き込みコマンドの) がキャッシュ・データ212 場合、ローカル・ストレージ106はステップ802、 リューム・タイプを調べる(ステップ801)。 特に、 る。スケジューテ220は、コンフィギュレーション ラム (サーバ214、215、又は216) を決定す (図2) 内に格納されているかいないかを調べる。 [0061] 特に、サーバ・プログラム213の適切な ージ106はステップ804を飛ばして、ステップ80 サーバ (214、215、又は216) は、書き込みコ マンドのデータがローカル・ディスク209内にあるか (書き込みコマンドで要求した) がローカル・ディスク 209内で割り当てられている場合、ローカル・ストレ 5に進む。ステップ803で、データがローカル・ディ スク209内にない場合、ローカル・ストレージ106 どうかを調べる。ステップ803で、すべてのデータ はステップ804を実行する。

カル・ストレージ106は、キャッシュ・ディレクトリ

210 (図2) を調べる (ステップ802)。

【0062】 (告き込みコマンドの) データがローカル ・ディスク209内で割り当てられていない場合、ロー カル・ディスク209とキャッシュ・データ212の両 に、キャッシュ制御218は、ローカル・ディスク20 9及びキャッシュ・データ212の両方でデータを割り カル・ストレージ106は書き込みデータについてロー 方にデータ領域を割り当てる (ステップ804)。 特

【0063】ローカル・ストレージ106は、キャッシ ディスク209に書き込む。特に、適切なサーバ・プロ 又はSCSIサーバ216)はデータをローカル・ディ ュ・データ212を介して書き込みデータをローカル・ グラム (NFSサーバ214、C1FSサーバ215、 スク209に書き込む。

ギュレーション・テーブル112(図4)を使用してデ ータを書き込む際のボリューム・タイプを調べる(ステ ップ806)。 特に、 サーバ・プログラム (NFSサー パ214、CIFSサーバ215、XはSCSIサーバ 【0064】ローカル・ストレージ106は、コンフィ

209にデータを格納し、このデータをプロバイダ・サ

イト (たとえば、サービス・プロバイダ・サイト10

216) はコンフィギュレーション・テーブル112を 使用してボリューム・タイプを調べる。ボリューム・タ

(図5) が設定されている場合、ローカル・ストレージ 【0066】続いてローカル・ストレージ106は、続 [0065] ステップ806で、ポリューム・タイプが 井同期領域の場合、ローカル・ストワージ106は保留 07)。特に、サーバ・プログラム(サーバ214、2 **降に、スケジューラ220は、統計情報113を更新す** 106はこの書き込みコマンドのログ・データをアクセ ス・ログ114に追加する。特に、スケジューラ220 データ・リスト211 (図2)を更新する (ステップ8 は書き込みコマンドのログ・データをアクセス・ログ1 計情報113(図5)を更新する(ステップ809)。 15、又は216)が保留データ・リストを更新する。 る。統計情報113内のログ・ポインタ (ログPtr) 14に追加する。これで書き込みプロセスは終了する。 [0067] 同期プロセス

信される。このスケジュールは、たとえば、図4のコン 込み)の場合、データは、ユーザ又はサービス・プロバ 非同期ボリューム書き込み(非同期領域へのデータ書き イダによって定められたスケジュールに従ってサービス ・プロバイダ・サイト(たとえば、サイト102)に法 エントリ内の値により定められる。スケジューラ220 (図2) は、シンクロナイザ217 (図2) を定期的に 実行することで、サービス・プロバイダ・サイトとデー タの同期をとることができる。この期間は、システム1 0.0ミリ秒に設定できる。図9は、本発明の実施形態 フィギュレーション・テーブル 1 1 2内の「ポリシー」 による同期プロセスの流れ図である。同期プロセスは、 00設定に応じて、たとえば、約1.0ミリ秒又は1 シンクロナイザ217 (図2) によって実行できる。

【0068】シンクロナイザ217は、保留データ・リ テップ901)。 最初に、シンクロナイザ217は、保 スト211 (図2) 内のヘッド・データを選択する (ス に) ある場合、シンクロナイザ217は以下で説明する 留データ・リスト211を調べる (ステップ902)。 保留データ・リスト211にデータがなければ、シンク ロナイザ217は同期プロセスを終了する。同期すべき データが1つ又は複数(保留データ・リスト211内 ようにステップ902~905を実行する。

特開2002-324000

ータ・リスト211内のこの保留データの同期をとる必 ジ106内のキャッシュ制御218は、コンフィギュレ ・プロバイダ・サイト (たとえば、サイト102) に格 **要があれば、シンクロナイザ217は保留データをサー に送り、保留データとサービス・プロバイダ・サイトと** D」及び「リモート・サイト」の値に基づいてサービス ビス・プロバイダ・サイト (たとえば、サイト102) の同期をとる(ステップ904)。ローカル・ストレー -ション・テーブル112 (図4) 内の「リモートI 笹中ベキゲータを認識する。

【0070】シンクロナイザ217は、保留データ・リ スト211内の次のデータを避択する (ステップ90

5)。シンクロナイザ217は、保留データ・リスト2 ~905を繰り返す。保留データ・リスト211に選択 するデータがなくなったら、図9の方法は終了する。コ ンピュータが上述のいずれかの方法を実行できるように 1 1 内で選択するデータがなくなるまでステップ902 コンピュータで読み取り可能な媒体に格納できるプログ ラム又はコードを実装することも本発明の範囲である。

【0071】そこで、本発明について特定の実施形態を 正、さまざまな変更、及び置き換えを許容することが意 図されており、規定されているとおり本発明の範囲を逸 脱することなく他の特徴の対応する使用がなくても場合 引用しながら説明してきたが、前記開示においては修

20

|図面の簡単な説明] められるであろう。

によっては本発明のいくつかの特徴を採用することは認

|図1||本発明の実施形態によるシステムのブロック

[図2] 図1に示されているローカル・ストレージの-**実施形態の詳細を示すブロック図**

జ

【図3】本発明の実施形態によるデータ・レイアウトの -例を示すブロック図。 [図4] 本発明の実施形態によるコンフィギュレーショ ン・ドーブルの一例を示すブロック図。

[図5]本発明の実施形態による統計情報の一例のブロ

[図6] 本発明の実施形態によるアクセス・ログの一例 ック 図 図

【図7】本発明の実施形態による読み込みプロセスの流 のブロック図

49

[図8] 本発明の実施形態による書き込みプロセスの流

[図9]本発明の実施形態による同期プロセスの流れ

[符号の説明]

101…ユーザ・サイト、102, 103…サービス・ プロバイダ・サイト、104, 105…ホスト、106 …ローカル・ストレージ、107…SAN、108…L **AN、111…仮想ストレージ**

20

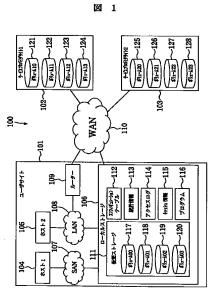
スト211の情報を調べる (ステップ903)。 保留デ

【0069】シンクロナイザ217は、保留データ・リ

(14)

特開2002-324000

[⊠]



N [図2] M

[図[]

9 M

5

略相	0:0:0	0:0:0	0:0:1	0:0:1	2:00	0.0.2	0:0:2									
B#	2000.8.1	2000.8.1	2000.8.1	2000.8.1	2000.8.1	2000.8.1	2000.8.1									
										•						
							-304		-116	-217	-218	233	-219			
	ĵ		ž,	_			<u> </u>	•	Ţ,	Τ,	Ĕ,	7	7]	608,	
	<u> </u>		- 1		135	~	キャッシュ金乗		70774	シンクロナイザ	キャッシュ信仰	2752-5	デバイスドライバ	١,	~~	٦
	40.5	3	4	2	~114		#		ا_ا	븯	#	<u>"</u>	Ä	$\prod_{i=1}^{n}$	8	
٦		ا	Ī	たようして	_		3	データリスト	イプログラム	NFS +-11	CIFS #-A	SCSI ∯−14			ローカルディスク	
	113	₩	۷,	į	クセスログ		Market States	1 3		Š	E	g			Î	

230

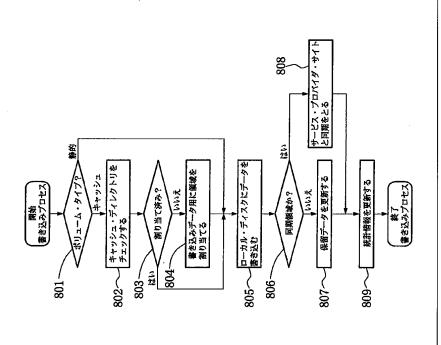
図 4 まリジー ¥ 図 3 コンフィギュレーションチーブル 8 2 2 £ [図4] サナッシュ ジモート (学区部) コピー (SEE SEE) None D インタ サブエリア リモート リモート ポリューム サーフェース サブエリア サイト ID タイプ がわが 8 [🖾3] 2 = 13 ZZ Bester (filt)
ZZ Bester (filt)
ZZ Hota str (filt)
ZZ Hota str (filt) 気器ストアージ OS CIFS 10 SCSI OS NFS 808

特開2002-324000

(12)

(17) 特開2002-324000





レロントページの流み

Fターム(参考) 58065 CE21 EA35 EK05 58082 FA11 FA12 GB06 HA02

(18)

特開2002-324000

[外国語明細書]

1 Title of Invention

VIRTUAL VOLUME STORACE

2 Claims

1. A system for providing a data storage service, the system comprising

a service provider site configured to provide a data storage service

a user site coupled by a wide area actwork (WAM) to the service prorider site, the user site comprising a local storage having a virtual storage, the virtual storage having a spuckroners volume and an asynchronous stoinme, the local storage configured to immediately transmit to the service provider site data that is written in the synchronous volume, to transmit at a predetermined schedule to the service provider site data that is written in the asynchronous volume, and to read data from the service provider site if the data is not stored in the local storage.

- The system of claim I wherein the local storage is configured to record an access activity to the local storage by the host.
- The system of claim I wherein the local storage records an access activity to a volume to the virtual storage.
- 4. The system of claim I wherein the local storage records an access activity in a sub area of a volume in the virtual storage.
- 5. The system of claim 2 wherein the access activity is provided to the service provider site for analysis.
- 6. A method of providing a data storage service, the method comprising:
 providing a local storage baving a virtual storage, the virtual storage comprising a synchronous volume; and

selectively performing at least one of the following:

(19)

site data that is written in the asynchronous volume the virtual storag transmitting at a predetermined schedule to the service provider

reading data from the service provider site if the data is not tored in the local storage.

- 7. The method of claim 6, further comprising:
- recording an access activity to the local storage.
- The method of claim ?, further comprising:

providing a record of the access activity to the service provider si

9. An electronically-readable medium storing a program for permitting a computer to perform a method of providing a data storage service, the m ethod comprising:

selectively performing at least one of the following:

immediately transmitting to a service provider site data that is written in a synchrenous volume of a virtual storage; transmitting at a predetermined schedule to the service provider site data that is written in an asynchronous volume the virtual storage reading data from the service provider site if the data is not s tored in a local storage that includes the virtual storage.

perform a method of providing a data storage service, the method compris 10. A program code embedded on a carrier wave for causing a computer to

selectively performing at least one of the following:

inmediately transmitting to a service provider site data that is written in a synchronous volume of a virtual storage;

site data that is written in an asynchronous volume the virtual storage transmitting at a predetermined schedule to the service provider

特開2002-324000

8

reading data from the service provider site if the data is not s ored in a local storage that includes the virtual storage. il. An apparatus for processing data in a data storage system, the appar tus comprising:

r site data that is written to the synchronous area, to transmit at a p a user site comprising a local storage having a virtual storage, the de local storage cenfigured to immediately transmit to a service provid to the asynchronous area, and to read data from the service provider sit redetermined schedule to the service provider site data that is written virtual storage including a synchronous area and an asynchronous area, e if the data is not stored in the local storage. 12. A method of processing data in a data storage system, the method com

in response to a command signal to a local storage, selectively perf orming at least one of the following: immediately transmitting to a service provider site data that is written to a synchrenous area in a virtual storage; transmitting at a predetermined schedule to the service provider site, data that is written to an asynchronous area in the virtual storag reading data from the service provider site if the data is not s tored in the local storage.

13. The method of reading data in a data storage system, the method comp

receiring a read command from a bost to a local storage;

determining the rolume type of the data that is requested by the rea

(22)

d command;

if the rolume type is static, then reading the data from a local dis k in the local storage; if the valume type is cached, then checking if the data is stored in the local disk;

if the data is stored in the local disk, then reading the data from the local disk; if the data is not stored in the local disk, then reading the data f rem a remote service provider site, storing the data in the local disk, and reading the data from the local disk;

after reading the data from the local disk, updating statistical inf ormation relating to the read command; and

returning the data to the host.

14. An apparatus for processing data in a data storage system, the appar atus comprising:

is written to a synchronous area in a rirtual storage, to receive at a p redetermined schedule data that is written to an aspachronous area in th 15. A method of processing data in a data storage system, the method com e virtual storage, and to transmit data to a local storage that includes a service provider site configured to immediately receive data that the virtual storage if the data is not stored in the local storage.

selectively performing one of the following:

immediately receiving data that is written to a, synchronous area in a virtual storage; receiving at a predetermined schedule data that is written to an asynchronous area in the rirtual storage; and transmitting data to a local storage that includes the virtual s

torage if the data is not stored in the local storage.

16. A method of writing data in a data storage system, the method compr

receiving a write command from a bost to a local storage;

determining the solume type of the data of the write command;

if the volume type is static, then writing the data to a local disk

in the local storage;

if the volume type is cached, then checking if the data is stored in the local disk; if the data is stored in the local disk, then writing the data to th local disk from a cache;

area in the local disk and in the cache and then writing the dats to the if the data is not stored in the local disk, then allocating a data local disk from the cathe;

checking if the volume type of the data is spuchronous;

if the volume type is synchronous, then immediately synchronizing th e date with a remote service provider site and then updating statistical information relating to the write command; and if the volume type is not synchronous, then synchronizing the data w ith a remote service provider site based on a predetermined schedule and 17. A system for providing a data storage service, the system comprising then updating statistical information relating to the write command.

means for selectirely performing at least one of the following:

immediately transmitting to a Service provider site data that is written in a synchronous volume of a virtual storage;

site data that is written in an asynchronous volume the virtnal storage transmitting at a predetermined schedule to the service provider

reading data from the service provider site if the data is not s

tored in a local storage that includes the virtual storage.

3 Detailed Description of Inrention

PIELD OF THE INVENTION

The present invention relates to data storage systems, and relates m ore particularly to a system and method for providing a virtual volume d ata storage in a user site.

BACKGROUND OF THE INVENTION

as providing storage volumes, data backup, and the like. From a user si s directory, then the user must use a long distance connection to connec t to this storage. This process causes more delays than if a user is co a data starage service provider provides a storage-related service such te, if a user connects to this storage via the storage service provider Managing a large data storage system is very difficult. Typically, naecting to a local storage. The World Wide Web (WWW) effectively acts as a storage system in a w een a client application (such as a web browser) and a remote server. T te servers. The proxy server intercepts all requests that are made to t d falfill the request. If the proxy server is unable to fulfill the req nest, then the proxy screer will forward the request to the remote serve ide area. The user may deploy a proxy server in the user site. As know a to those skilled in the art, a proxy server is a server that sits betw he proxy server provides a cache of items that are available on the remo he remote server so that the proxy server can determine if it can instea r for processing. A proxy server can just only read cache data from the remote serrer or site, and it does not support a write procedure to the remote serrer or site. Additionally, the proxy server can not provide a service that depends on the data feature. As a result, the proxy serv

er causes an ineffective usage of resources, and the use of a proxy sorr

特開2002-324000

(24)

er also causes more delays than the use of a local storage.

ose starage systems between wide area networks (WANs) and ignores data f pose (user data or system file or backup file and the like). Thus, the s estures such as access interface type (block or file, and the like), pur U.S. patent 5,155,835 discloses a multilerel, bicrarchical, dysamica Iy napped data storage subsystem. This patent reference does not disch ibsystem disclosed in this reference is inefficient. The luternet protocol small computer system interface (iSCSI) uses t ing protocols, iSCSI (IP SAN) can unify network architectore across an e ntire enterprise, thereby reducing the overall network cost and complexi and utilities that have been developed for 1P networks. The iSCSI prot ocal is discussed, for example, at the website, http://www.ece.com.edu/ he Interact Protocol (1P) networking infrastructure to quickly transport large amounts of block storage (SCSI) data over existing local area and for wide area networks. With the potential to support all major network ty. To easure reliability, iSCSI can use known actwork management tools ips/index.btml) in the IP Storage section, http://www.ece.cmu.edu/Tips/D ce/doce.html.

of IETF) and a common Internet File system (CIFS) protocol (http://www.c The working grapp of Internet Engineering Tank Porce (IETF) at hith: //www. ieif.org provides a network file system (NFS) version 3 (rici813 ifs.org).

PS provides access to shared files through an interface called Virtual P access shared files that are stored in different types of computers. N ile System (VPS) which runs on top of the Transmission Control Protocul/ The NFS is an open operating system that allows all, network neers to internet Protocol (TCP/1P). With NFS, compaters connected to a network can operate as clients while accessing remote files and as servers while

providing remote users access to local shared files.

The CIFS protocol defines a standard for remote file access using mi llions of computers at a time. With CIFS, users with different platform CIFS rans over TCP/1P, but uses the Server Message Block (SMB) protocol found in Microsoft Windows for file and printer access. Therefore, CIF Swill allow all applications (including Web browsers) to open and share s and computers can share files without having to install new software. files across the Internet. The websites bittp: //www.cnt.com and bitp: //www.san.com describe wi (SANE). The WAN connection may be an asynchronous transfer mode (ATM), s de area network (WAN) connections to a storage or storage area networks yachronous optical network (SONET), and the like.

The above references do not disclose methods for providing bieraschi cal management techniques between storage systems and host devices.

lso a need for a system and method that will permit an access service pr -mentioned deficiencies of conventional methods and systems. There is a There is a need for a system and method that will overcome the above quickly access the storage system. There is also a need for a system an ovider to bare a reliable storage system and that will permit a user to d method that will permit an access service provider to be able to tune a local storage system based upon user statistic data and user log data.

usly provide a virtual local storage that permits faster access for a us er. The present invention may also advantageously permit a service prov local storage in a user site. The present invention may also advantageo ider to tune up the local storage by using user statistics data and user data that are tracked by the local storage, and the service provide The present invention may advantageously provide a reliable virtual

(36)

特開2002-324000

r may then charge the user for these tuning services.

geously permits the oser to aroid baying to access each line the wide a ome storage is deployed at the user site. The virtual volume storage a located at the service provider site. The virtual volume storage advant ca actwork coupled between the user site and the service provider site. o provide a data storage service via a wide area network. A virtual vo llows a bost(s) at the user site to access the service provider storage In one embodiment, the present invention permits a service provider

In one embodiment, the present invention broadly provides a system f or providing a data storage service, comprising: a service provider site configured to provide a data storage service; and a user site coupled b comprising a local storage having a virtual storage, the wirtual storage having a synchrondes relame and an asynchronous volume, the local slora d schedule to the service provider site data that is written in the asyn that is written in the synchronous volume, to transmit at a predetermine " a wide area network (WAM) to the strvice provider site, the user site ge configured to immediately transmit to the service provider site data chronous volume, and to read data from the service provider site if the data is not stored in the local, storage.

naging data in a local storage based on the data feature. The virtual v olume storage in the local storage determines if the data is static or c In another embodiment, the present invention provides a method of ma also manage data not only by storage volome, but also by directory, file ached, and synchronous or asynchronous. The virtual volume storage can , cylinder, and/or block address. In another embodiment, the present invention provides a method of tr cing the access activities of a user in the user site. The user access activities are recorded to permit the service provider to analyse patte ns in the user access activities. Based on this analysis, the service

provider can tune the virtual volume storage in the oser site.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The following description is provided to enable any person skilled i ications to the embodiments will be readily apparent to those skilled in the art, and the generic principles defined berein may be applied to ot her embodiments and applications without departing from the spirit and s cope of the present invention. Thus, the present invention is not inten ded to be limited to the embodiments shown, but is to be accorded the wi context of a particular application and its requirements. Various medif dest scope consistent with the principles, features, and teachings disc n the art to make and use the present invention, and is provided in the

In one embodiment in accordance with the present invention, a system 100 is configured by a user site 101, and by a first service provider s site 101 is described below. The number of service provider sites in t ite 102 and a second service provider site 103. The details of the user he system 100 may vary. The first service provider site 102 includes st site 103 includes storage volumes 125, 126, 127, and 128. The number of der sites 102 and 103 may be located in, for example, two different safe buildings. The user site 101, first service provider site 102, and sec Stornge volunes in a service provider site may rary. The service provi The service providers 102 and 163 provide data storage servi orage volumes 121, 122, 123, and 124, while the second service provider and service provider site 103 are each connected to a Wide Area Network ces ria the WAN 110 to a user at the user site 101. (WAN) 110.

As known to those skilled in the art, a WAN is a computer network th at typically spans a relatively large geographical area. A WAN may also

(82)

特開2002-324000

ten consected through public networks, such as the telephone system. T include local area networks (LANs). Computers connected to a WAN are c hey can also be connected through leased lines or satellites. The large st WAN in existence is the Internet. The WAN 110 permits communication between each site (user site 101 a ly cenfigured by asynchronous transfer node (ATM), synchronous optical n itwork (SONET), Deuse Waveleagth Division Maltiplexing (DWDM), or Intern ld service provider sites 102 and 103). Each site 101, 102, and 103 mag be separated from each other by long distances. The WAN 110 is typical et Protocol (IP) network.

The hosts 104 and 105 access the local storage 106. The hosts 104 and 1 nay wary. The LAN 108 is connected to the WAN 110 wia, for example, a r outer 109. Prom the teachings of the present invention berein, it is un derstood by those skilled in the agt that the user site 101 may be imple mented to include only one of the SAN 107 or LAN 108, or both the SAN 10 At the user gite 101, a first bost 104 is connected to a local stora ge 106 via a storage area actwork (SAN) 107, while a second bost 105 is 05 are, for example, servers. The number of hosts in the user site 101 consected to the local storage 106 via a local area actwork (LAN) 148. 7 and LAN 108.

As known to those skilled in the art, a SAN is a high-speed sub-netw As more storage devices are a ork of shared storage devices wherein the SAN makes all storage devices oded to a SAN, these added storage devices will also becacessible from available to all servers in a LAN or WAN. ing server in the larger network.

The SAN 107 may be configured with, for example, fibre chainel or Sm art, a fibre channel is a serial data transfer architecture dereloped by all Computer System interface (SCSI). As known to those skilled in the a consortium of computer, and wass storage device manufacturers and now

being standardized by the American National Standards Institute (ANSI).

peripheral devices that require a very high bandwidth. Using optical fi berg to connect the devices. PC-AL supports fall-duplex data transfer ra The most prominent fibre channel standard is the Fibre Channel Arbitrat ed Loop (FC-AL) which is designed for new mass storage devices and other tes of approximately 100 megabytes per second (MBps). As also known to those skilled in the art, SCSI is a parallel interf ace standard used by Apple Macintosh computers; personal computers (PCs) SCSI interfaces provide for faster data transmission rates (up to about , and many UNIX systems for attaching peripheral devices to computers. 80 MBps) than standard serial and parallel ports.

but it is also able to access data and devices anywhere on the LAN. Th ns, many users can bhare expensive devices, such as laser printers, as w by, for example, sending e-mail or engaging in chat sessions. There as PCs.. LANs are capable of transmitting data at very fast rates, much fa to a single building or group of buildings. Most LANs connect workstati ons and personal computers. Each node (individual computer) in a LAN ha e many different types of LANs, with Ethernets being the most common for ster than the data transmitted over a telephone line. Howerer, the dist As also known to those skilled in the art, a LAN is a computer netwo s its own central processing unit (CPU) with which it executes programs, ell as data. Users can also use the LAN to communicate with each other, rk that typically spans a relatively small area. Nost LANs are confined ances over LANs are limited, and there is also a limit,on the number of computers that can be attached to a single LAN.

local storage 106 by. for example, a file (NFS or CIFS) I/F. The second In Pigure 1, the LAN 108 is configured as, for example, an Ethernet. The first host 104 accesses the local stornge 106 by use of, for exa mple, a block (SCSI) interface (1/P). The second host 105 accesses the

lost 185 may also access the local storage 106 by, for crample, a block

特開2002-324000

8

(iscsi) 1/F.

The local storage 105 and service provider sites 102 and 103 may use ISCSI or fiber channel over ATM, fiber channel over SOMET, or a maique ender protecel.

re 5), access log 114 (see also Figure 6), cache information 115, and p y vary. These volumes 117-120 are managed by the local storage 106. Th e local storage 106 and service provider sites 102 and 103 collaborate f able 112 (see also Figure 4), statistical information 113 (see also Fig regrams 116. The components 112, 113, 114, 115, and 116 permit the loca storage 106 to act like a carbe. The hosts 104 and 105 can access the rirtual storage 111. The virtual storage 111 includes some volumes 117 118, 119 and 120. The number of volumes in the virtual storage 111 ma The local storage 106 includes a virtual storage 111, configuration or creation of the virtual volume storage [11].

1865) at the user site 101. The virtual volume storage 111 allows the u eers of host 104 and/or bost 10% to access the hage service provider sto ge ill is deplayed at the oser site 101. The virtual valume starage li (volumes 121-124 and/or volumes 125-128). The virtual volume storage 1 1/F (NPS, C1PS) as discussed below. The virtual volume storage 111 adv antiageously permits the user to aroid in baring to access the WAN 110 er rage (rolumes 121-124 and/or volumes 125-128) as the virtual volume stor I typically has a smaller size than the service provides storage volumes [] has several interfaces such as, for example, a block 1/F (SCS1), file ery time. The virtual volume storage 111 also enables the service provi The user can deploy the virtual velume storage 111 (in local storage der to provide a faster and more reliable storage system for the user.

Figure 2 is a black diagram showing the details of one embodiment of

(31)

特開2002-324000

the local storage 106. The local storage 106 includes a microprocessor 201 (which is available from, for example, latel Corporation or Motoroll a Corporation), a local (internal) has 202, a SAN interface (1/F) 203 caupled to the SAN 107 (Figure 1), a LAN 1/F 204 coupled to the LAN 108 (Figure 1), a memory 207 campled to the WAN 110 (Figure 1), a memory 1/F 205 coupled to the WAN 110 (Figure 1), a memory 1/F 206, a memory 207 contains the configuration table 112, statistical information 1 13, access 10g 114, cache information 115, and programs 116. The microprocessor 201 controls all resources in the local storage 106 and execute all procedures in the local storage 106 and execute all procedures in the local storage 116.

Figure 2 shows a snapshot of the local disk 209 during the running of procedures in the local storage 106. As known to those skilled in the art, a snapshot is a copy of a whele rolame or a part of a rolume of a real storage. These information and programs shown in Figure 2 are stored in the local disk 209. At the boot sequence of local storage 106, the ese data and programs are mored from the local disk 209 to the memory 20

The cache information 115 is configured with cache directory 210 is a directory 210 is a directory (configuration information) of the cache directory 210 is a directory (configuration information) of the cached data 212 in memory 20 T and local disk 209. This cache data 212 may be managed by the least recently used (LRU) management. But some cases should be an exception of the LRU management, because there will be no possibility to access again in the near future. One such exception is sequential, access where the storage address is accessed in a sequential manage.

The pending data list 211 is a list of pending data which is saved in the local storage 106. The pending data list 211 has information of each pending data such as a pointer to pending data, a pointer to a service provider site, and a spechronous period. A spackronous period is def

ined by the data feature. If the data is not important, then this data may be stored in the local storage 106 for a predetermined period. This period may be, (or example, one minute, one hour, one day, one week, or one month. If the data is important, then after the data is stored in the virtual storage 111, the data should be stored immediately (synchron ons) in a service provided site (e.g., site 102) by using the configuration table 112 (Figure 4). For example, synchronous data is stored in Volume 00 117. In Figure 4, Volume 00 117 is identified with the identific cation number 110 00. Since the data stored in Volume 00 117 is synchron ons data, the data will be immediately stored in the service provider site 102 (Remote 51 te 1) at Volume 10 121 (Remote 10 10).

This synchraneus access feature provides a reliable data storage system. However, syrchronous access requires much time because of the access time by the oser site 101 to the service provider sites 102 and/ar 103. The required access time depends on the particular operation being carried out on the system 100.

Data that are accessed by hosts 104 and/or 105 are stored in the memory 207. This data is called cache data 212. Cache data 212 is a tempo rary data that is frequently accessed by a user of host 104 and/or host

To one embodiment, the programs 116 include server programs 213, a synchronizer 217, a cache control 218, a scheduler 220, and a device drive er 219. The server programs 213 permit the local storage 106 and the host illust (and/or host 105) to communicate with each other. The server programs 213 are configured with an NPS server 214, a CIPS server 215, and an SCSI server 216. The NPS server 214 acts as a conventional WPS server 115 and are 126 server 215 and as SCSI server 215 acts as a conventional CIPS server 215 and are 126 acts as a conventional CIPS server 125 acts as a conventional CIPS server 125 acts as a conventional CIPS server 126 acts as a SCSI target device. The synchronizer 217 synchronizes the local storage 106 and the service provider sites 102 ani/ar 103 by

using the pending data list 211. The cache control 218 controls cache 2
12 in the memory 207 and local disk 209 by using the cache directory 210
. The scheduler 220 schedules all procedures such as the serrer programs 213, spackronizer 217, cache control 218, and device driver 219. The device driver 219 controls all devices in local storage 106 such as the SAN I/F 203, the LAN I/F 204, the WAN I/F 205, the memory 1/F 206, and the edisk controller 208.

Pigare 3 is a block diagram showing one example of a data layout in accordance with an embodiment of the present invention. The management of the local storage 106 is based upon the use of data feature, as described below in additional details. The wirtual volume storage 111 deals with the data feature, depending on whether the data is static or cacked, synchronous or asynchronous. As also described ferther below, the rirtual volume storage 111 deals with data feature not only by storage volumes, but also by directory file, cylinder, and/or block address. At user site 101, there are four (4) types of virtual volumes (Volume 00 1 17, Volume 01 118, Volume 02 119, and Volume 03 120) in the virtual storage 111. The features of these viginal volumes are described bellow: (3) the static synchronous volume, (2) the static asynchronous volume, (3) the cacked asynchronous volume.

(1) The static synchronous volume, and (4) the cacked asynchronous volume.

Volume 00 117 is a static synchronous volume. Thus, Volume 10 117 of compies the same size as the real storage volume, Volume 10 121, in the first service provider site 102. The hosts 184 and/or 105 (at aser site 101) read the data from Volume 00 117 and not from Volume 10 121. If the accessed data is in the cache data cache 212 (Figure 2), then the local storage 106 (Figure 2) does not access the local disk 209 (Figure 2)

and just accesses the memory 207. The bests 104 and/or 105 write data to both Velume 00 117 and Volume 10 121. This static synchronous relume, Volume 00 117, provides a fast read access and reliable storage system.

(2) The static asynchronous Volume (Volume 01 118):

特開2002-324000

(34)

Volume 01 118 is a static asynchronous volume. Thus, Volume 001 118 occupies the same size as the real storage volume, Volume 11 122, in the first service provider site 102. The hosts 104 and/or 105 read data from just only Volume 01 118 and not from Volume 11 122. If the accessed data is in the cache data 212, then local storage 106 doce not access the elocal disk 209 and only accesses the nemory 207. The hosts 104 and/or 105 write data to Volume 01 118 and register to the pending data list 2 if (Figure 2). This pending data is synchronized with a predetermined schedule, as described below.

This pending data may be compressed in the background job in the local storage 106 (Figure 2) to reduce the WAN 110 traffic. This static as yachronous volume, Volume 01 118, pravides a fast read and write volume, but may not provide the same reliability as a synchronous volume.

(3) The cached spachronous volume (Yelume 02 119):

Volume 02 119 is a cached syachronous volume. Thus, Volume 02 119 or comples same size as the real storage volume, Yolome 12 123, in the first service prorider site 102. Volume 02 119 acts as a cache. Thus, there is an eatle in Yolume 02 119 that is accessed by the bosts 104 and/or 105. If the data is not in the local disk 209 (Figure 2), then the local storage 106 (Figure 2) change 105 and values the read data to the local disk 209. After the data is in the local storage 106 sends the data is in the local disk 209. After the data is in the local storage 106 sends the data from the local disk 209. If the accesse cal disk 209 via memory 207 to the hosts 104 and/or 105. If the accesse

or 105 write data to Volume 02 119 and Volume 12 123. This cacked synch roncus volume, Volume P2 119, advantageously permits the reduction of si re of the local storage 106. This virtual volume, Volume 02 119, also p The bests 104 and/ d data is in the cache data 212, then the local storage 196 does not acc ess local disk 209 and just accesses the menory 207. rovides a reliable storage system.

(4) The cached asynchronous volume (Volume 03 120);

occupies same size as the real storage volume, Volume 13 124, in the fir st service provider site 10%. Volume 04 1:20 acts as a cache. Thus ther e is no data in Veltme 03 120 that is accessed by the bosts 104 and/or 1 05. If the data is not in the local disk 209, the local storage 106 rea ds data from the first service provider site 102 and writes this read da ta into the local disk 209. After the data is already in the local stor age 106, the local storage 106 sends the data from local disk 209 via me mory 207 to the hosts 184 and/or 185. If the accessed data is in the da and just accesses the memory 207. The hosts 104 and/or 105 write the da ta to Volume 01 118 and register to pending data list 211 (Figure 2). T his pending data is spackronized with a schedule. This cached aspuchron ta cache 207, then the local storage 106 does not access local disk 209 ous volume. Volume 03 120, advantageously permits the reduction of size of the local storage 106, but may not provide the same reliability as a Volume 03 120 is a erched asgnehrongus volume. Thus, Volume 03 120 synchronous volume.

123 is duplicated on rolling 22 127 in the second service provider site i Pigure 3 also illustrates a service provided by the, service provider site. Volume 11 122 in the first service provider site 102 is duplicat Volume 13 124 is duplicated on volume 23 128 in the second service ed on volume 21 125 in the second service provider site 183.

(36)

特開2002-324000

ionally, at the first service provider site 182, a backup of the data i Volume 12 123 may be made (see arrow 132) without requiring the functi The volume data may be duplicated by use of, for example, conventional remote mirroring technology. Addi rorider site 103. This duplication (remote copying 130) of volume data on of the user site 101. This backup service permits the user to reduce provides a disaster recovery method. the workload of data back up. As now discussed below, the virtual volume storage 111 can make a tr ice data based upon access activities from the bosts 104 and/or 105. Th e virtual volume storage III can provide to the service provider this tr ace data for purposes of analyzing the user access pattern. After analy ing the user access pattern, the service provider can advantageously tu ne the virtual volume storage 111.

nterface", "Sub Area", "Renote Site", "Remote 10", "Volume Type", "Servi this case in the example of Figure 4, each sub area has the same volume type. For example, Volume 00 117 (Figure 3) (with ID 00) is a static sy schronous volume type as mentioned previously. Remote Site indicates an Figure 4 is a block diagram of an example of a configuration table 1 12 in accordance with an embodiment of the present invention. The confi juration table 112 is configured with the following parameters: "ID", "i ce" ("Backup", 'Analyze"), "size" ("Total", "Used", "Free"), and "Policy itorage 196. For cxample, Volune 80 117 (Figure 3) has an ID 66. Inter JPS.. Sub area indicates the number of sub areas in the virtual volume. A sub area may be a certain area of block address, cylinders, file, an for directory. Volume type may be different between egen sub area. In identifier of the remote site (e.g., the identifier of service provider . ID indicates volume identification of a virtual volume in the local lace indicates the virtual volume interface type such as SCSI, NPS. and

特開2002-324000

dicates volume 1D in a service provider site. For example, Volume 10 12 Valume type indicate s a type of rolome such as static or cached, synchronous or asynchronous Service indicates a service provided by the service provider Such as a back up service or an analyze service. When a service provider provide s an analyze service, the service provider sets a pointer to a statistic al information 113 (Figure 2) and creates statistical information (acces s log) il4 to obtain a data feature. If an analyze service is provided, then "Ptr" will be indicated in the Analyre section in the Configuratio n Table 112 as shown in the example of Figure 4. For example, Volume 02 119 (ID 02) in Configuration Table 112 has a pointer Ptr in the Analyze section (see row 400). This pointer for ID 02 (of Volume 02 119) point s to 1D 02 in the Statistical Information 113 in Pigure 5. 1D 02 has a Sub area (directory) "/asra" (see row 500) with a pointer ("ptr") in the Log Ptr section. This pointer (ptr) points to the access log 114 which site 102). This may be the IP address of an http address. Remote ID in is an access log created for Volume 02 119 in the example of Pigure 6. i in service provider site 102 has a remote [D 10.

It is further noted that in the example of Figure 5, the 1D 02 (of V oleme 02 119) includes a directory "/usrh" (see row 505) with a "Nall" y alse (as pointer) in the Log Ptr section. Additionally, 10 02 includes a directory "/usrc" (see row 510) with a Null value (no pointer) in the Log Pir section. Thus, so associated access logs 114 bare been created for the directories /usrb and /usrc since these directories do not hare pointers that point to an access log.

is not in the file system in local storage 106, then only the total size needs to be known. The parameter policy indicates a schedule for synch In the Configuration Table 112, Size indicates total memory size (To ronization. For example, in row 410, the Yolume 00 117 (with ID 00) has tal) in gigabytes, used size (Used), and free size (Free). If a volume

is synchronized daily with data in a service provider site (c.g., site 1 Volume 02 119 (with ID 02) has a "week" value set in the Po site 102). Valume 03 120 (with ID 03) has a "day" value set in the Po alue indicates that a synchronization schedule has not been set for Vol 's synchronised once per week with data in a service provider site (e.g. licy parameter, and this ralue indicates that the data in Volume 02 119 icy parameter, and this value indicates that the data in Volume 03 120 a "nult" value set in the Policy parameter in Figurer 4, and this null une 00 117.

The parameters in the Configuration Table 112 are set by sciting the configuration data in the Configuration Table 112. Each storage subsyst em has an associated Configuration Table. Pigure 5 is a block diagram showing an example of statistical inform ation 113 in accordance with an embodiment of the present invention. Wh enabled. The local storage 106 collects the user statistical access inf ormation of local storage 106 in this table of Statistical Information 1 13. The Statistical Information is configured with "ID", "Sub Area", "I O/s" and "MB/s" (Read Are, Max, Write Are, Max), "Hit Ratid", and "Log P en a user ar service provider sets the Service-Analyze parameter to "Ptr tr". The parameters ID and Sub area (in Figure 5) baye the same meanings as ID and Sub area is Configuration Table 112 (Figure 4). 19/s indicate is in a volume directory may be recorded by use of any suitable known me in the Configuration Table 112 (Figure 4), then the following will be s the number of host/aser input or output activities per second such as read and write commands per eccond. A read or write agcess in a valone

Hects each read and write occurrence in a volume (or in a sub area in a MB/s indicates number of bytes per second. The local storage 106 co

so tracks the maximum occurrence (Max) of read and write occurrence in a The local storage 106 volume), and averages (Ave) this occurrences. volume or in a sub area in a volume.

Hit Ratio indicates a hit ratio of local storage. This bit ratio is , for example, only just for the local disk 209 (Figure 2). A bit ratio is defined in equation (1):

(1) Hit ratio = 100%(#LDCAL DISK NEAD/#ALL READS)

where #LOCAL DISK READ is the number of reads in the local disk 209 by a host, and #ALL READS is the number of total reads by the bost. Statistical Information 113 is sent to the service provider sites to 2 and/or 103 periodically. The service provider analyzes the feature of data, and from this analysis the service provider can propose better so lations to the user of a host at the user site 101.

If the service provider wants to know more detail of user access patter as, then the service provider can get this pointer to a created accesses As stated above, Log Pir indicates a pointer to the access log 114. 10g 114.

e date of an access by a user. Time indicates the time of an access. C ommand indicates the command type (e.g., read or write) of an access. F vice provider wants to analyze the details of a user's access paitern, t Log Ptr (in Statistical Information 113 in Figure 5) to this access log 114. The local storage 106 collects an access log 114-of each access by a user. Ibia acress lag ilf is configured with the parameters: "Date", "Time", "Command", "File ID", "Address", and "Size". Date indicates th in accordance with an embodiment of the present invention. When the ser hen the service provider creates an access ing 114 and sets a pointer at Figure 6 is a block diagram showing an example of an access log 114 lle 1D indicates the identification of a file that was accessed by this

(40)

梅開2002-324000

command. If access log is for SCSI, the File, ID may be aull. Address a nd size indicates the access address and size.

ata (which is requested by the read command) is in the cache data 212, t command. Each server program 213 checks whether the data requested by t he read command is in or not in the cache data 212 (Pigure 2). If all d by the read command) is not in cache data 212, then the requested data s cache. The hosts 104 and/or 105 access the local storage 106, since th e local storage 106 has a large space. At first, a bost (e.g., host 104 In the read procedure, the cached device (local storage 106) acts as AM 108 (Figure 1), or a read command of SCS1 protocol via SAN 107. The 203 (Pigure 2). The device driver 219 (Figure 2) bandles this read comma scheduler 220 then analyzes the read command and places the read comman or 185) via internal bus 202 and a network interface (SAN 1/F 203 or LAN hould, be mored to cache data 212 from the local disk 209 or from the ser or host 105) issues a read command of NFS, CIFS, or iSCSI protocol via 1/F 204). On the other hand, if all of or part of the data (requested local storage 106 receives the read command via LAN 1/F 204 or SAN 1/F i into a proper server (NFS server 214, CIFS server 215, or SCSI server 216) within the server programs 213 for purposes of processing the read hen the local storage 106 returns data to the requesting bost (bost 104 nd from a bost and places this read command into the scheduler 220. vice provider sites 102 or 103.

storage 186 checks 701 for the volume type of the data. The scheduler 2 Figure 7 is a flowchart diagram of this read process in accordance w ith an embodiment of the present invention. After the local storage 106 28 checks configuration table 112 for the volume type (e.g., static type receires a read command from a host (e.g., bost 104 or 105), the local

特開2002-324000

The local georage 106 checks 703 the cacke directory 210. In particular, the appropriate serrer (214, 215, or 216) in the serrer program 21 3 checks whether the data requested by the read command is in or not in the local disk 209. If all data (requested by the read command) is stored in the lacal disk 209, then the local storage 106 skips step 704 and proceeds to step 705 which is described below. If data (request by the read command) is not in the local disk 209, then local storage 106 executes step 704.

disk 200 and reads data (requested by the read command) from a service provider site (e.g., service provider site 102). In particular, the cach e control 218 uses the configuration table 112 (Figure 4) to determine the service provider site should be obtained. The cache control 218 reads the data from where data should be obtained. The cache data is the local disk 209.

After data is stored in local disk 209, the local storage 106 will more (read) 705 the data from the local disk 209 to the cache data 212. In particular, the cache control 218 mores the data from the local disk 209 to the cache data 212. The local storage 106 them updates 706 the s

tatistical information [13 (Figure 5). In particular, the scheduler 220 in the local storage 106 updates the statistical information [13. [f. t be log painter (Log Ptr) (Figure 5) in the statistical information [13.] set, then the local storage 106 adds the log data of this read command to the access log 114 (Figure 5). The scheduler 220 or the device driver 219 creates the statistical information [13 and access [og 114. Typical], the scheduler 220 is preferred in creating the statistical information [13] and access log 114.

The data (requested by the read command) is then mored 707 from the cache data 212 to the host (e.g., host 104 or 105) that sent the read command. In particular, the cache control 218 mores the data from the cac be data 212 to the appropriate server program (214, 215, or 216), and the eappropriate server program (214, 215, or 216) returns the data to the host that sent the read command.

Write process

In the write procedure, the cached device (local storage 106) acts as a cache. The host 104 and/or host 105 access the local storage 106, s ince the local storage 106, s ince the local storage 106 has a Targe space. At first, a host (e.g., host 104 or host 105) issues a write command of NCSI protocol via SAN 107. The local storage 106 receives the write command via LAN 1/F 204 or SAN 107. The local storage 106 receives the write command via LAN 1/F 204 or SAN 1/F 203. The device driver 219 (Figure 2) handles this read command from a host and places this read command into the scheduler 220 then analyzes the write command and places the write command into a proper server (NRS server 214, CIFS 215, or SCSI server 218) within the server programs 213. Each server program 213 checks whether the data of the write command is in or not in the cache data 212 (Figure 2). If all data (of the write command) is in the cache data 212, then the local s

d, then the local storage 106 stores data in that allocated area. When a li over write the data on an allocated space in the local disk 209 and c allocate an area (in local disk 209 and cache data 212) for the rest of li data is stored in the cache data 212, then the local storage 106 stor twork interface (SAN 1/F 203 or LAN 1/F 204). The local storage' 106 wi acte 212. On the other hand, if all of or part of the data (of the Writ e command) is not in the cache data 212, then the local storage 106 will data not in the cache data 212. After all of the data area is allocate torage 106 receives the data from the host via internal bus 202 and a ae es data in local disk 209 and sends this data to a provider site (e.g., service provider site 101).

e is a static valume, then the local storage 106 skips steps 802, 803, a han embodiment of the present invention. After the local storzge 196 r eceires a write command from a bost, the local storage 106 checks 801 th or 216) should process (bandle) the read command. The Scheduler 226 lo oks at configuration table 112 (Wigure 4) for the volume type and the ap propriate serrer program to handle the write command. If the volume typ nd 804, and proceeds to step 805 which is described below. If the rolum e type is a cached volume, then the local storage 106 zhecks if the data e volume type of the data of the write command. In particular, the sche type of the data and to determine which server program (server 214, 215. (of the write command) is stored or not stored in cache data 212 (Figur Figure 8 is a flowchart diagram of a write process in accordance wit deler 220 (Figure 1) analyzes the write command to determine the volume

program 213 checks whether the data of the write command is allocated i In particular, the appropriate server (214, 215, or 216) in the server n the local disk 209. In step 803, if all the data (requested by the wr The local storage 196 checks 802 the cache directory 210 (Figure 2).

(44)

特開2002-324000

106 skips step 804 and proceeds to step 805. In step 803, if the data i ite command) is allocated in the local disk 209, then the local storage s not in the local disk 209, then local storage 106 executes step 804.

cache control 218 aliocates the data area on both the local disk 209 and 209, then the local storage 106 allocates 804 data area in both the lo If the data (of the write command) is not allocated in the local dis cal disk 209 and cache data 212 for the write data. In particular, the cache data 212.

ogram (NFS server 214, CIFS server 215, or SCSI server 216) writes the d sk 209 via the cache data 212. In particular, the appropriate serrer pr The local storage 106 them writes 805 the write data to the local di ate to the local disk 209.

the volume type is a synchronous area, then the local storage 106 immedi ately spechronizes (writes) 808 the data to a service provider site (e.g ser site) to the service provider site. The cache control 218 in the l The local storage 106 checks 806 the volume type in which the data i ., site 102) and proceeds to stpp 809 which is discussed below. The syn ocal storage 105 knows where the data should be stored in the service pr s written by use of the configuration table 112 (Figure 4). In particul 216) oses the configuration table 112 to check for the volume type. If thronizer 217 (Figure 2) performs the synchronizing of the data (in the ar, the server program (MFS server 214, CIPS server 215, or SCSI server ovider site by use of the configuration table 112.

In step 866, if the roltme type is an asynchronous area, then the lo ticular, the server program (server 214, 215, ar 216) updates the pendin tal storage 106 updates 807 the pending data list 211 (Rigure 2). In par

The local storage 106 then updates 809 the statistical information 1 (Figure 5). In particular, the scheduler 220 updates the statistical

Synchronization Process

In the case of an asynchronous volume write (a data write to an asynchronous area), the date should be sent to a service provider site (e.g., site 102) by a schedule that is defined by the user or service provide.

I. This schedule is, for example, defined by the value in the "Pelicy" entry in the configuration table 112 in Figure 4. The scheduler 220 (Figure 2) executes the synchronization with a service provider site. This period may be set to, for example, approximately 10 mili-second or 10.0 mili-seconds, depending on the system 100 settings.

Figure 9 is a flowchard diagram of synchronization process in accordance with an embodiment of the present invention. The synchronization process may be performed by the synchronizer 217 (Pigure 2).

The synchronizer 217 selects 901 the head data in the pending data ist 211 (Figure 2). At the first, the synchronizer 217 checks 902 the pending data list 211. If there is no data in the pending data list 211, there is no data in the pending data list 211, then the synchronization process. If there is one or more data (in the pending data list 211) which should be synchronized, then the synchronizer 217 executes steps 902-905 as described below.

The synchronizer 217 checks 903 the information of pending-data list 211. If this pending data in the pending data list 211 should be synchronized, then the synchronizer 217 sends the pending data to a service p

(46)

梅開2002-324000

nowider site (e.g., site 102), so that the pending data is synchronized 904 with the service provider site. The cache control 218 in the local storage 106 knows where data should be started in the service provider site (e.g., site 162) based upon the values in the Remote (D and Remote Site in the configuration table 112 (Figure 4).

The synchronizer 217 then selects 905 the next data in the pending data list 211. The synchronizer 217 repeats steps 902 through 905 until there is no more data to select in the pending data list 211. If there is no more data to select in the pending data list 211, then the method of Figure 9 eads.

It is also within the scope of the present invention to implement a program or code that can be stored in an electronically-readable medium to permit a computer to perform any of the methods described abore.

Thus, while the present invention has been described berein with reference to particular embodiments thereof, a fatitude of modification, various changes and substitutions are intended in the foregoing disclosure, and it will be appreciated that in some instances some features of the invention will be employed without a corresponding use of other features swithout departing from the scope of the invention as set forth.

4 Brief Description of Brawings

Figure 1 is block diagram of a system in accordance with an embodine of the present invention;

Figure 2 is a black diagram showing additional defails of one embodinger of the local storage in Figure 1;

Figure 3 is a block diagram showing one example of a data layout in accordance with an embodiment of the present invention;

(48)

Figure 4 is a block diagram of an example of a configuration table in a accordance with an embodiment of the present invention;

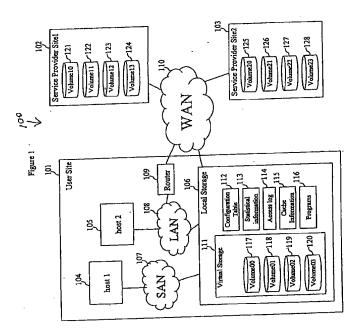
Figure 5 is a black diagram of an example of statistical information in accordance with an embodiment of the present invention;

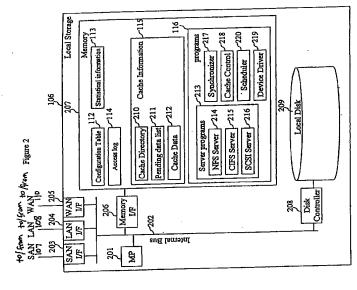
Figure 6 is a block diagram of an example of an access log in accord ance with an embodinent of the present invention:

Figure 7 is a flowchart diagram of a read process in accordance with an embediment of the present invention;

Figure 8 is a flowchart diagram of a write process in accordance with an embodineal of the present invention; and

Figure 9 is a flowchart diagram of synchronization process in accord ance with an embodiment of the present inreation.





特開2002-324000

(49)

Figure 5

Statistical information

				8.	_	1	ļ.	ŧ	\$0 \$0\$
513		ğ ,	Ę		Noil	il.	FINE.	Mul	
	-				Š				
		۰	ž	Г	2.	Ĺ.,			
	3/	Write	γvc		0.02				
	MB/s	Read	Max		ន				
		2	Ave		0.12				•••
		2	Max		ฆ				
		Write	γĸ		3				
	IO/s	Read	XXIA		হ				
			Ave		13				
		Sub Area				z.isn/	/usrb	/userc	
		A		00	10		8		

Figure 6

1000	3	0x0001	0×0100	0x0100	0x0100	0x0100	0x0100		
Address Size	0x0000	0×0000	0x0100 0	0×0200	0×0200 0	0x0300	0x0400 0		
File ID	0	0	10	10	10	10	10		
Command	Read	Write	Read	Read	Read	Read	Read		
Time	0:0:0	0:0:0	0:0:1	0:0:1	0:0:5	0:0:5	0:0:5	! !	
Date	2000.8.1	2000.8.1	2000.8.1	2000.8.1	2000.8.1	2000.8.1	2000.8.1	_	

n I

Null 20 20 Ptr 100 50

12

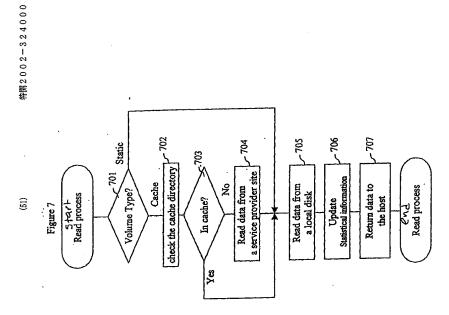
02 NFS 3

01 SCSI None 00 SCSI None

Configuration Table

Figure 4

Figure 3 102 Real Storage 103	Nolume 10 125 Nolume 20 125 Nolume 20 125 Nolume 21 Nolume 21	Service Provider site: Service Provider site: 10.2
	Vértual Storage	nchronous) ynchronous) nretronous) synchronous)



特開2002-324000

特開2002-324000

(23)

provider site configured to provide a dala storage service; and a user 8 ite couplei by a wide area network (WAN) to the service provider site, t he user site comprising a local storage baying a virtual storago. the vi rtual storage baring a synchronous rolume and an asynchronous volume, th e local storage configured to immediately transmit to the service provid er site data that is written in the synchronous volume, to transmit at a predetermined schedule to the service provider site data that is wrilte n in the asynchronous rolume, and to read data from the service provider A system for providing a data storage service, comprises: a service site if the data is not stored in the local storage.

. 902 No

ending Data?

7901

Select pending data

Synchronize process

Figure 9

2. Representative Drawing

7 98

select next data

Synchronize with Service provider site

Should be

ž

8nchronize process